

# Rangkuman Bimbingan Teknis Penulisan Skripsi Bidang Minat Data Mining dan Computer Vision

**Muhammad Jauhar Vikri, M.Kom**

# Muhammad Jauhar Vikri

- NIDN : **0712078803**
- S1 Teknik Informatika (S.Kom)
  - Intelligent System & Computer Vision.
  - Image Recognition
  - Image Detection
  - Image Enhancement
- S2 Teknik Informatika (M.Kom)
  - Intelligent System
  - Data Mining & Machine Learning
  - Deep Learning

# Pembahasan

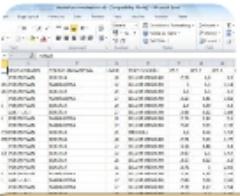
1. Bimbingan Skripsi
2. Topik Penelitian Data Mining dan Business Intelligence
3. Pengantar Penelitian/Skripsi
4. Tahapan Penelitian
5. Literature Review
6. Menulis Skripsi

# Bimbingan Skripsi

- Sebaiknya **bimbingan tatap muka** / hadir dikampus, namun jika **tidak bisa hadir tetap harus melaporkan progress via link Gform.**
- **Semua Konsultasi** dan Pertanyaan Skripsi Wajib disampaikan melalui **Grup WA**.
- **Wajib Ada progres setiap minggu,**
  - progres perkembangan tulisan/laporan,
  - progress software/aplikasi
  - progress pembelajaran atau temuan masalah-masalah serta kesulitan dan usaha penyelesaian.
- Form Perkembangan [Progres](#)

# Data Mining & Business Intelligence

# Proses Data Mining



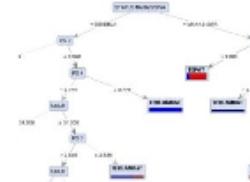
## 1. Himpunan Data

(Pahami dan Persiapkan Data)

$$\int_{-\infty}^{\infty} f(x) dx = \sqrt{2\pi} \left( e^{-x^2/2} \sum_{n=0}^{\infty} \frac{x^n}{n!} \left( \frac{e^{-x^2}}{2^n} \right)^n \right)$$
$$= \left( -\sigma \varphi(\frac{x}{\sigma}) \right) \left| \frac{\partial}{\partial x} \right| + \frac{\sigma^2}{2} \left[ \ln(1+2\sigma^2) + \frac{r}{\sigma^2} \text{Erf}(2\sigma) \right]$$
$$= \sigma \varphi \left( \frac{x}{\sigma} \right) \left( 1 + \frac{1}{2} \sigma^2 \right) \left( \ln(1+2\sigma^2) + \frac{r}{\sigma^2} \text{Erf}(2\sigma) \right)$$
$$w_p = \left[ \frac{f_p(x)}{f_p(x)} \right] = \left[ \frac{1}{\sigma^2} \right] \left[ \ln(1+2\sigma^2) + \frac{r}{\sigma^2} \text{Erf}(2\sigma) \right]$$

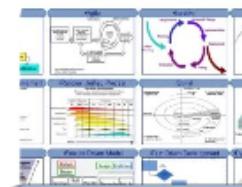
## 2. Metode Data Mining

(Pilih Metode Sesuai Karakter Data)



## 3. Pengetahuan

(Pahami Model dan Pengetahuan yg Sesuai )



## 4. Evaluation

(Analisis Model dan Kinerja Metode)

**DATA PREPROCESSING**  
*Data Cleaning  
Data Integration  
Data Reduction  
Data Transformation*

**MODELING**  
*Estimation  
Prediction  
Classification  
Clustering  
Association*

**MODEL**  
*Formula  
Tree  
Cluster  
Rule  
Correlation*

**KINERJA**  
*Akurasi  
Tingkat Error  
Jumlah Cluster*

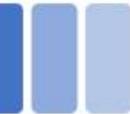
**MODEL**  
*Atribut/Faktor  
Korelasi  
Bobot*



# Data Cleaning

**Data in the Real World Is Dirty:** Lots of potentially incorrect data, e.g., instrument faulty, human or computer error, transmission error

- **Incomplete:** lacking attribute values, lacking certain attributes of interest, or containing only aggregate data
  - e.g., Occupation="" (missing data)
- **Noisy:** containing noise, errors, or outliers
  - e.g., Salary="-10" (an error)
- **Inconsistent:** containing discrepancies in codes or names
  - e.g., Age="42", Birthday="03/07/2010"
  - Was rating "1, 2, 3", now rating "A, B, C"
- Discrepancy between **duplicate records**
  - Intentional (e.g., disguised missing data)
  - Jan. 1 as everyone's birthday?



# Incomplete (Missing) Data

- Data is **not always available**
  - E.g., many tuples have no recorded value for several attributes, such as customer income in sales data
- **Missing data** may be due to
  - equipment malfunction
  - inconsistent with other recorded data and thus deleted
  - data not entered due to misunderstanding
  - certain data may not be considered important at the time of entry
  - not register history or changes of the data
- Missing data may need to be inferred

# Contoh Missing Data

- Dataset: **MissingDataSet.csv**

ExampleSet (11 examples, 0 special attributes, 15 regular attributes)															View Filter (11 / 11): all		
Row No.	Gender	Race	Birth_Year	Marital_Status	Years_on_Inf	Hours_Per_Day	Preferred_Browser	Preferred_Search	Preferred_Email	Read_News	Online_Shopping	Online_Gaming	Facebook	Twitter	Other_Social		
1	M	White	1972	M	8	1	Firefox	Google	Yahoo	Y	N	N	Y	N	?		
2	M	Hispanic	1981	S	14	2	Chrome	Google	Hotmail	Y	N	N	Y	N	?		
3	F	African Amer	1977	S	6	2	Firefox	Yahoo	Yahoo	Y	Y	?	Y	N	?		
4	F	White	1961	D	8	6	Firefox	Google	Hotmail	N	Y	N	N	Y	?		
5	M	White	1954	M	2	3	Internet Expl	Bing	Hotmail	Y	Y	N	Y	N	?		
6	M	African Amer	1982	D	15	4	Internet Expl	Google	Yahoo	Y	N	Y	N	N	?		
7	M	African Amer	1981	D	11	2	Firefox	Google	Yahoo	?	Y	?	Y	Y	LinkedIn		
8	M	White	1977	S	3	3	Internet Expl	Yahoo	Yahoo	Y	?	?	Y	99	LinkedIn		
9	F	African Amer	1969	M	6	2	Firefox	Google	Gmail	N	Y	N	N	N	?		
10	M	White	1987	S	12	1	Safari	Yahoo	Yahoo	Y	?	Y	Y	N	MySpace		
11	F	Hispanic	1959	D	12	5	Chrome	Google	Gmail	Y	N	N	Y	N	Google+		



# Data Reduction Methods

- Data Reduction
  - Obtain a reduced representation of the data set that is much smaller in volume but yet produces the same analytical results
- Why Data Reduction?
  - A database/data warehouse may store terabytes of data
  - Complex data analysis take a very long time to run on the complete dataset
- Data Reduction Methods
  1. Dimensionality Reduction
    1. Feature Extraction
    2. Feature Selection
      1. Filter Approach
      2. Wrapper Approach
      3. Embedded Approach
  2. Numerosity Reduction (Data Reduction)
    - Regression and Log-Linear Models
    - Histograms, clustering, sampling



# 1. Dimensionality Reduction

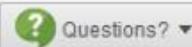
- Curse of dimensionality
  - When dimensionality increases, data becomes increasingly sparse
  - Density and distance between points, which is critical to clustering, outlier analysis, becomes less meaningful
  - The possible combinations of subspaces will grow exponentially
- Dimensionality reduction
  - Avoid the curse of dimensionality
  - Help eliminate irrelevant features and reduce noise
  - Reduce time and space required in data mining
  - Allow easier visualization
- Dimensionality Reduction Methods:
  1. Feature Extraction: Wavelet transforms, Principal Component Analysis (PCA)
  2. Feature Selection: Filter, Wrapper, Embedded



Views:

Design

Results



### Repository

Add Data

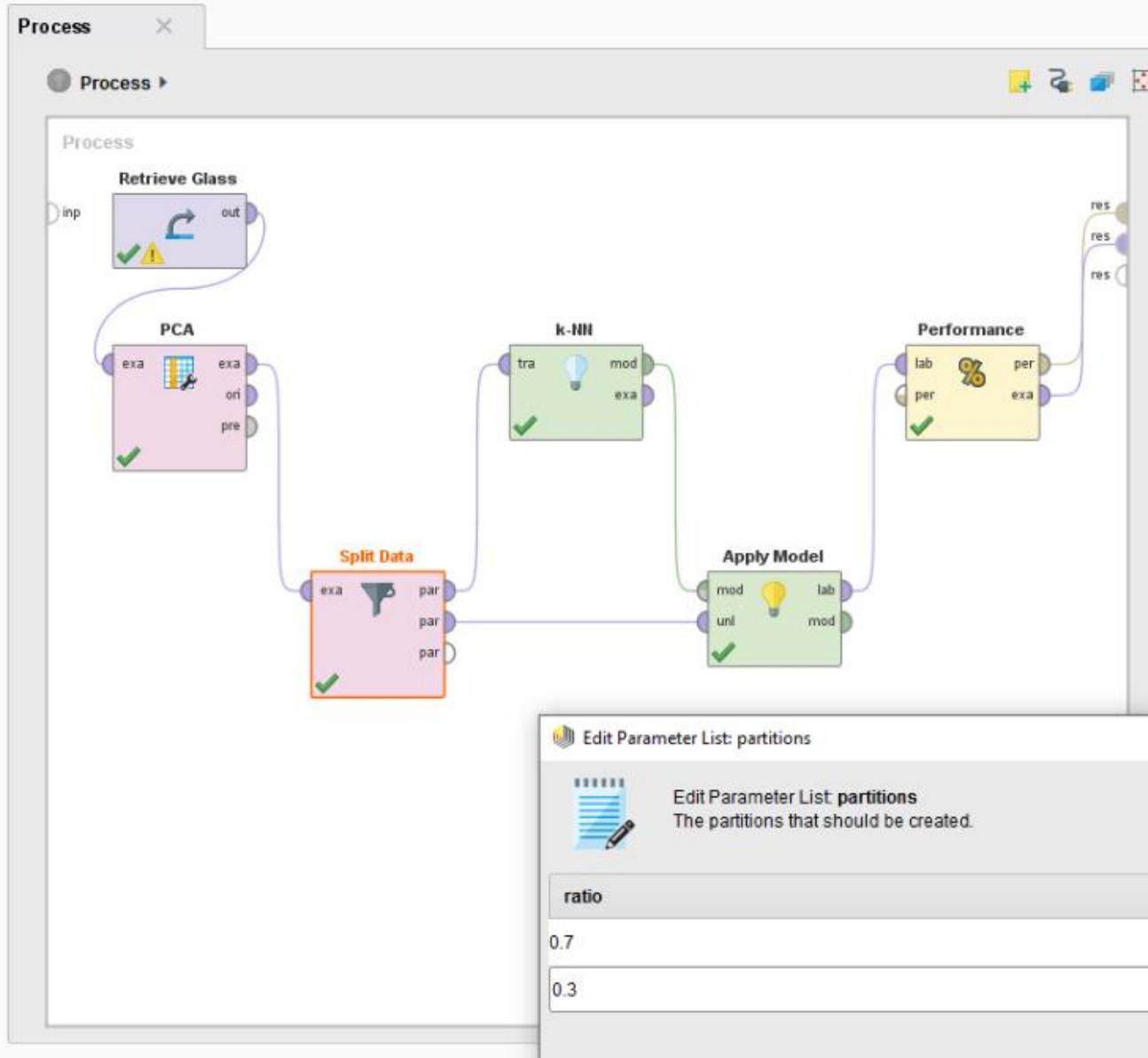
- HargaSaham (RomiSatria - v1)
- HeatingOil (RomiSatria - v1)
- HeatingOil-Scoring (RomiSatria - v1)
- IMFCountry (RomiSatria - v1)
- MusicGenre (RomiSatria - v1)
- SportSkill (RomiSatria - v1, 2)
- SportSkill-Scoring (RomiSatria - v1)
- Transaksi (RomiSatria - v1)
- MissingValueData (RomiSatria - v1)
- Glass (RomiSatria - v1, 2/25)

### Operators

#### Data Editor

Row No.	Id (integer) id
1	1
2	2
3	3
4	4
5	5
6	6

Repository Location: //Local R...



### Parameters

Split Data

partitions

sampling ... strati... ▾

use local random se

Hide advanced

## Edit Parameter List: partitions



Edit Parameter List **partitions**  
The partitions that should be created.

ratio

0.7

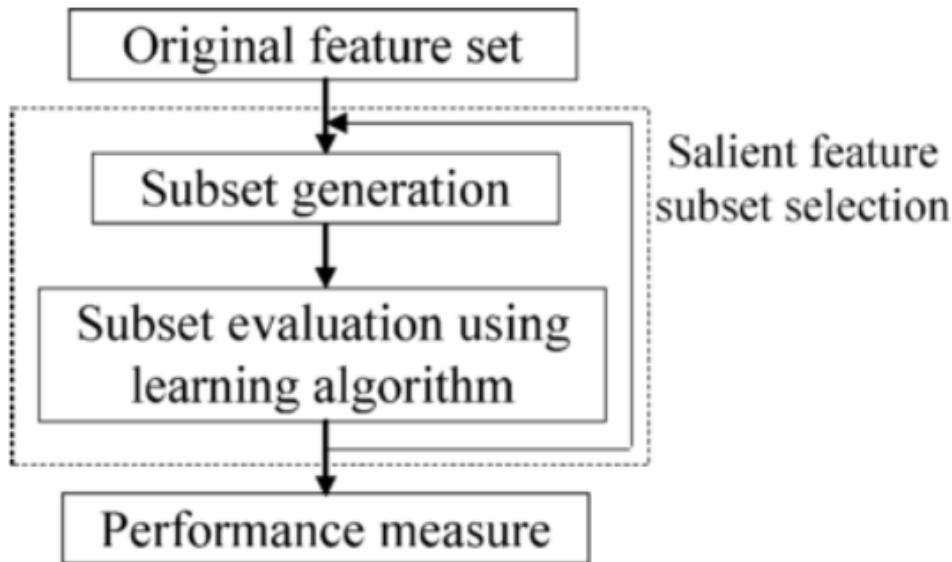
0.3

# Feature Selection Approach

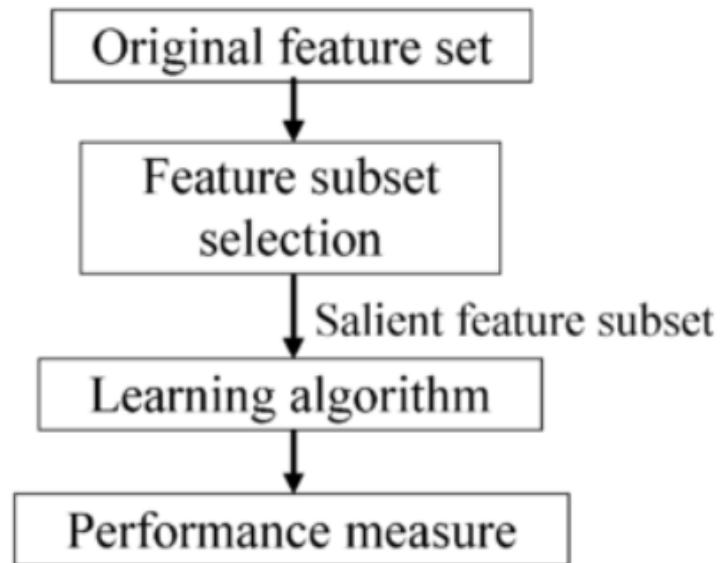
A number of proposed approaches for feature selection can broadly be categorized into the following three classifications: **wrapper**, **filter**, and **embedded** (Liu & Tu, 2004)

1. In the **filter approach**, statistical analysis of the feature set is required, **without utilizing any learning model** (Dash & Liu, 1997)
2. In the **wrapper approach**, a predetermined learning model is assumed, wherein **features are selected that justify the learning performance** of the particular learning model (Guyon & Elisseeff, 2003)
3. The **embedded approach** attempts to utilize the complementary strengths of the wrapper and filter approaches (Huang, Cai, & Xu, 2007)

# Wrapper Approach vs Filter Approach



Wrapper Approach



Filter Approach

# Feature Selection Approach

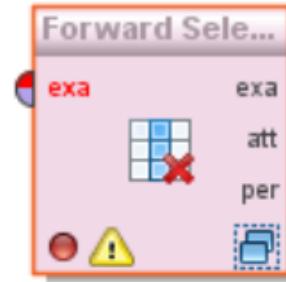
## 1. Filter Approach:

- information gain
- chi square
- log likelihood ratio
- etc



## 2. Wrapper Approach:

- forward selection
- backward elimination
- randomized hill climbing
- etc



## 3. Embedded Approach:

- decision tree
- weighted naïve bayes
- etc

# Latihan: Prediksi Kelulusan Mahasiswa

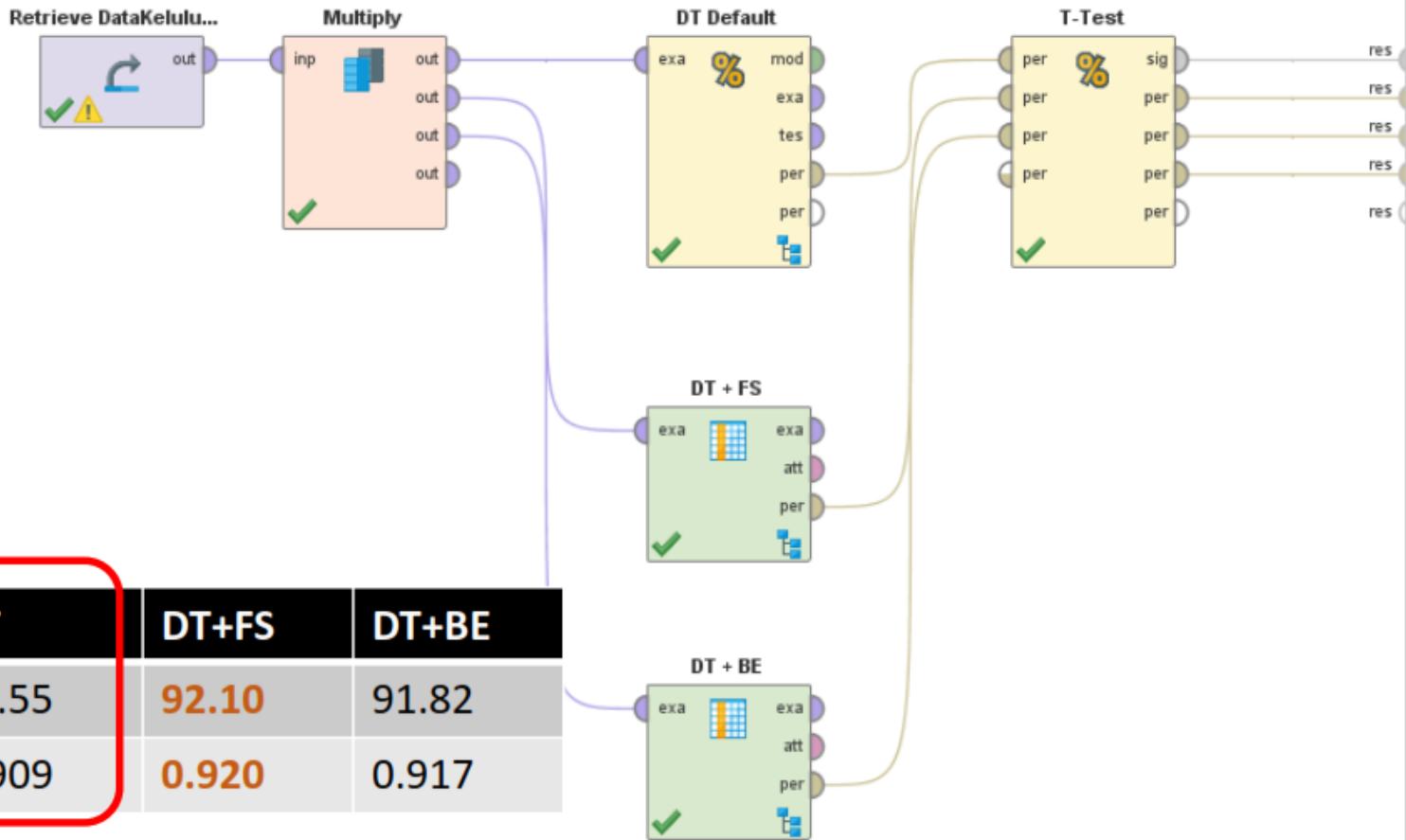
1. Lakukan **training** pada data mahasiswa ([datakelulusanmahasiswa.xls](#)) dengan menggunakan 3 algoritma klasifikasi (**DT, NB, k-NN**)
2. Analisis dan komparasi, mana **algoritma klasifikasi** yang menghasilkan **model paling akurat** (**AK**)
3. Lakukan feature selection dengan **Information Gain (Filter)**, **Forward Selection**, **Backward Elimination (Wrapper)** untuk model yang paling akurat
4. Analisis dan komparasi, mana **algoritma feature selection** yang menghasilkan **model paling akurat**
5. Lakukan **pengujian** dengan menggunakan 10-fold X Validation

	AK	AK+IG	AK+FS	AK+BE
Accuracy	91.55		<b>92.10</b>	91.82
AUC	0.909		<b>0.920</b>	0.917

# Latihan: Prediksi Kelulusan Mahasiswa

1. Lakukan **training** pada data mahasiswa ([datakelulusanmahasiswa.xls](#)) dengan menggunakan 4 algoritma klasifikasi (DT)
2. Lakukan feature selection dengan **Forward Selection** untuk algoritma DT (DT+FS)
3. Lakukan feature selection dengan **Backward Elimination** untuk algoritma DT (DT+BE)
4. Lakukan **pengujian** dengan menggunakan 10-fold X Validation
5. Uji beda dengan **t-Test** untuk mendapatkan model terbaik (DT vs DT+FS vs DT+BE)

	DT	DT+FS	DT+BE
Accuracy	91.55	<b>92.10</b>	91.82
AUC	0.909	<b>0.920</b>	0.917



The table compares the performance of four models (A, B, C, D) across three metrics: Accuracy, AUC, and F1 Score. Model A has the highest accuracy (0.916 +/- 0.039) and AUC (0.921 +/- 0.050). Model B has the highest F1 Score (0.787). Model C has the highest AUC (0.918 +/- 0.032). Model D has the highest F1 Score (0.867). The last three rows of the table are highlighted with red boxes, and a red dashed arrow points from the 'DT' row in the previous table to the 'B' row in this table, indicating a comparison between the two.

A	B	C	D
0.916 +/- 0.039	0.916 +/- 0.039	0.921 +/- 0.050	0.918 +/- 0.032
0.921 +/- 0.050		0.787	0.867
0.918 +/- 0.032			0.884

**no significant difference**



# Metode Data Mining

## 1. Estimation (Estimasi):

Linear Regression (LR), Neural Network (NN), Deep Learning (DL),  
Support Vector Machine (SVM), Generalized Linear Model (GLM), etc

## 2. Forecasting (Prediksi/Peramalan):

Linear Regression (LR), Neural Network (NN), Deep Learning (DL),  
Support Vector Machine (SVM), Generalized Linear Model (GLM), etc

## 3. Classification (Klasifikasi):

Decision Tree (CART, ID3, C4.5, Credal DT, Credal C4.5, Adaptative  
Credal C4.5), Naive Bayes (NB), K-Nearest Neighbor (kNN), Linear  
Discriminant Analysis (LDA), Logistic Regression (LogR), etc

## 4. Clustering (Klastering):

K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means  
(FCM), etc

## 5. Association (Asosiasi):

FP-Growth, A Priori, Coefficient of Correlation, Chi Square, etc



# Evaluasi Model Data Mining

## 1. Estimation:

- Error: Root Mean Square Error (RMSE), MSE, MAPE, etc

## 2. Prediction/**Forecasting** (Prediksi/Peramalan):

- Error: Root Mean Square Error (RMSE) , MSE, MAPE, etc

## 3. Classification:

- Confusion Matrix: Accuracy
- ROC Curve: Area Under Curve (AUC)

## 4. Clustering:

- Internal Evaluation: Davies–Bouldin index, Dunn index,
- External Evaluation: Rand measure, F-measure, Jaccard index, Fowlkes–Mallows index, Confusion matrix

## 5. Association:

- Lift Charts: Lift Ratio
- Precision and Recall (F-measure)



# Contoh Penerapan Kasus



Next Page .....



# 1. Business Understanding

- **Problems:**
  - Budi adalah Rektor di Universitas Suka Belajar
  - Universitas Suka Belajar memiliki masalah besar karena **rasio kelulusan mahasiswa tiap angkatan sangat rendah**
  - Budi ingin memahami dan membuat pola dari profile mahasiswa yang bisa lulus tepat waktu dan yang tidak lulus tepat waktu
  - Dengan pola tersebut, Budi bisa melakukan konseling, terapi, dan memberi peringatan dini kepada mahasiswa kemungkinan tidak lulus tepat waktu untuk memperbaiki diri, sehingga akhirnya bisa lulus tepat waktu
- **Objective:**
  - Menemukan pola dari mahasiswa yang lulus tepat waktu dan tidak

## 2. Data Understanding

- Untuk menyelesaikan masalah, Budi mengambil data dari sistem informasi akademik di universitasnya
- Data-data dikumpulkan dari data profil mahasiswa dan indeks prestasi semester mahasiswa, dengan atribut seperti di bawah
  1. NAMA
  2. JENIS KELAMIN: Laki-Laki atau Perempuan
  3. STATUS MAHASISWA: Mahasiswa atau Bekerja
  4. UMUR:
  5. STATUS NIKAH: Menikah atau Belum Menikah
  6. IPS 1: Indeks Prestasi Semester 1
  7. IPS 2: Indeks Prestasi Semester 1
  8. IPS 3: Indeks Prestasi Semester 1
  9. IPS 4: Indeks Prestasi Semester 1
  10. IPS 5: Indeks Prestasi Semester 1
  11. IPS 6: Indeks Prestasi Semester 1
  12. IPS 7: Indeks Prestasi Semester 1
  13. IPS 8: Indeks Prestasi Semester 1
  14. IPK: Indeks Prestasi Kumulatif
  15. STATUS KELULUSAN: Terlambat atau Tepat

### 3. Data Preparation

Data set: **datakelulusanmahasiswa.xls**

Row No.	STATUS KEL...	NAMA	JENIS KELA...	STATUS MA...	UMUR	STATUS NIK...	IPS 1	IPS 2
1	TERLAMBAT	ANIK WIDAYA...	PEREMPUAN	BEKERJA	28	BELUM MENI...	2.760	2.800
2	TERLAMBAT	DWI HESTYN...	PEREMPUAN	MAHASISWA	32	BELUM MENI...	3	3.300
3	TERLAMBAT	MURYA ARIE...	PEREMPUAN	BEKERJA	29	BELUM MENI...	3.500	3.300
4	TERLAMBAT	NANIK SUSA...	PEREMPUAN	MAHASISWA	27	BELUM MENI...	3.170	3.410
5	TERLAMBAT	RIFKA ISTIQF...	PEREMPUAN	BEKERJA	29	BELUM MENI...	2.900	2.890
6	TERLAMBAT	SUHARYONO	LAKI - LAKI	BEKERJA	27	BELUM MENI...	2.950	2.820
7	TEPAT	FARIKHATUN...	PEREMPUAN	MAHASISWA	26	BELUM MENI...	2.760	3.140
8	TEPAT	FIFI SUNALISA	PEREMPUAN	MAHASISWA	27	BELUM MENI...	2.620	2.890
9	TERLAMBAT	HENDRIK M...	PEREMPUAN	BEKERJA	25	MENIKAH	3.600	3.540
10	TERLAMBAT	IMAM AGUNG...	PEREMPUAN	BEKERJA	28	BELUM MENI...	2.710	2.550
11	TERLAMBAT	IMAM SANTO...	PEREMPUAN	BEKERJA	27	BELUM MENI...	3.140	3.460
12	TERLAMBAT	IRFAN EKO ...	PEREMPUAN	BEKERJA	32	BELUM MENI...	2.670	2.300
13	TERLAMBAT	IWAN HAMBALI	PEREMPUAN	BEKERJA	26	BELUM MENI...	2.570	2.820
14	TERLAMBAT	M SYAIFULLAH	PEREMPUAN	BEKERJA	31	BELUM MENI...	2.710	3

### 3. Data Preparation

- Terdapat 379 data mahasiswa dengan 15 atribut
- Missing Value sebanyak 10 data, dan tidak terdapat data noise

Name	Type	Missing	Statist...	Filter (15 / 15 attributes):	Search for Attributes
IPS 8	Real	7	Min 0	Max 4	
IPK	Real	3	Min 0.870	Max 3.850	
<b>STATUS KELULUSAN</b>	Binominal	0	Least TERLAMBAT (163)	Most TEPAT (216)	
NAMA	Polynominal	0	Least ZUMROTUN HALIMAH (1)	Most SRI LESTARI (2)	
JENIS KELAMIN	Binominal	0	Least PEREMPUAN (145)	Most LAKI - LAKI (234)	
STATUS MAHASISWA	Binominal	0	Least BEKERJA (133)	Most MAHASISWA (246)	
UMUR	Integer	0	Min 22	Max 50	

### 3. Data Preparation

- Missing Value dipecahkan dengan menambahkan data dengan nilai rata-rata
- Hasilnya adalah data bersih tanpa missing value

Retrieve data kelulusan...

Replace Missing Values

Name	Type	Missing	Statist...	Filter (15 / 15 attributes):
Label STATUS KELULUSAN	Binominal	0	Least TERLAMBAT (163)	Most TEPA
exa	Polynomial	0	Least ZUMROTUN HALIMAH (1)	Most SRI LE
ori	Binominal	0	Least PEREMPUAN (145)	Most LAKI -
pre	Binominal	0	Least BEKERJA (133)	Most MAHA
UMUR	Integer	0	Min 22	Max 50
STATUS NIKAH	Binominal	0	Least MENIKAH (8)	Most BELUM
IPS 1	Real	0	Min 0.330	Max 3.790

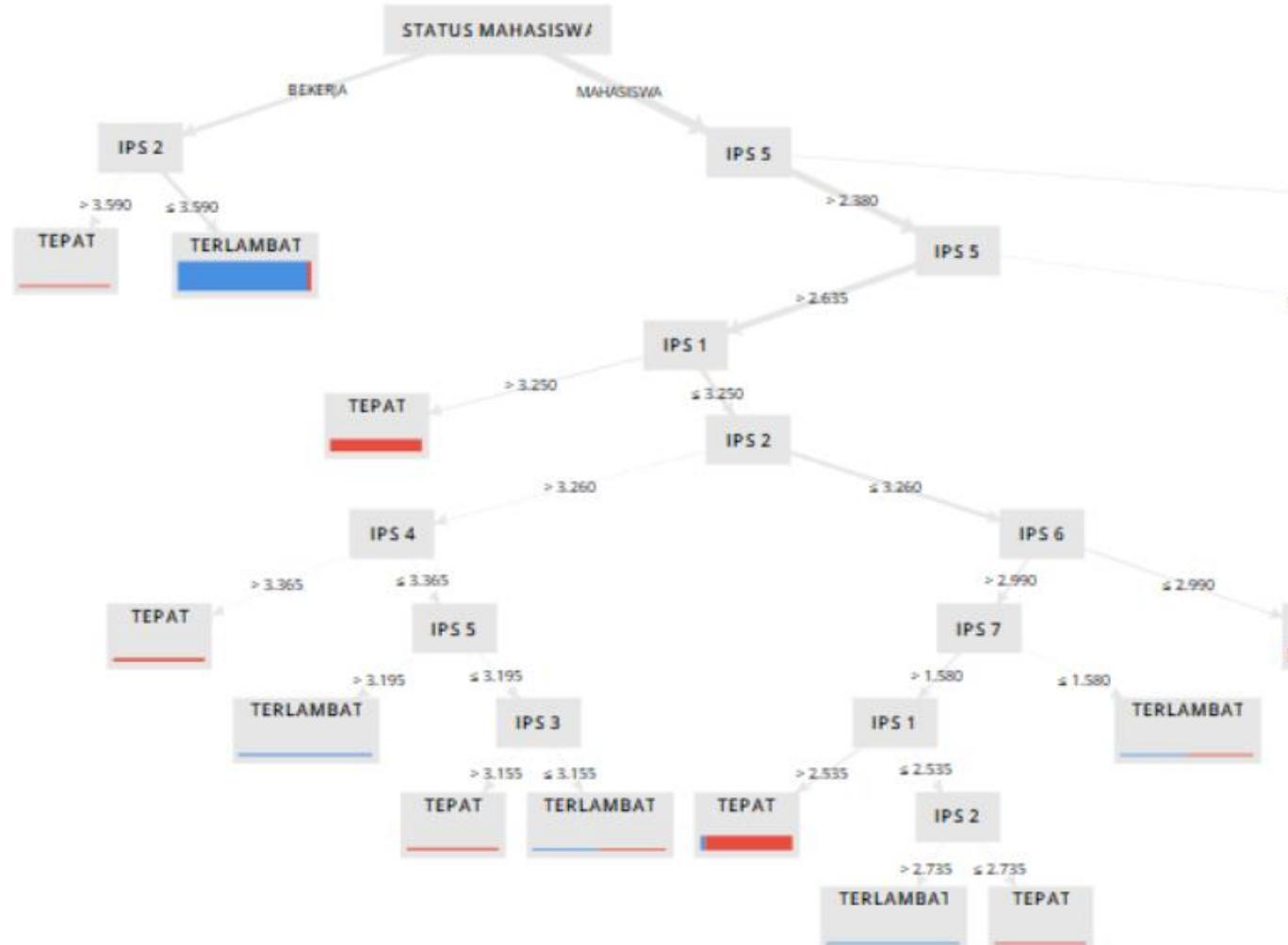
## 4. Modeling

- Modelkan dataset dengan Decision Tree
- Pola yang dihasilkan bisa berbentuk tree atau if-then



# 4. Modeling

Hasil pola dari data berupa berupa **decision tree** (pohon keputusan)



# 5. Evaluation

Hasil pola dari data berupa berupa peraturan if-then

```
STATUS MAHASISWA = BEKERJA
|   IPS 2 > 3.590: TEPAT {TERLAMBAT=0, TEPAT=2}
|   IPS 2 ≤ 3.590: TERLAMBAT {TERLAMBAT=127, TEPAT=4}
STATUS MAHASISWA = MAHASISWA
|   IPS 5 > 2.380
|   |   IPS 5 > 2.635
|   |   |   IPS 1 > 3.250: TEPAT {TERLAMBAT=0, TEPAT=50}
|   |   |   IPS 1 ≤ 3.250
|   |   |   IPS 2 > 3.260
|   |   |   |   IPS 4 > 3.365: TEPAT {TERLAMBAT=0, TEPAT=10}
|   |   |   |   IPS 4 ≤ 3.365
|   |   |   |   |   IPS 5 > 3.195: TERLAMBAT {TERLAMBAT=4, TEPAT=0}
|   |   |   |   |   IPS 5 ≤ 3.195
|   |   |   |   |   |   IPS 3 > 3.155: TEPAT {TERLAMBAT=0, TEPAT=5}
|   |   |   |   |   |   IPS 3 ≤ 3.155: TERLAMBAT {TERLAMBAT=1, TEPAT=1}
|   |   |   |   IPS 2 ≤ 3.260
|   |   |   |   IPS 6 > 2.990
|   |   |   |   |   IPS 7 > 1.580
|   |   |   |   |   |   IPS 1 > 2.535: TEPAT {TERLAMBAT=3, TEPAT=58}
|   |   |   |   |   |   IPS 1 ≤ 2.535
|   |   |   |   |   |   |   IPS 2 > 2.735: TERLAMBAT {TERLAMBAT=2, TEPAT=0}
|   |   |   |   |   |   |   IPS 2 ≤ 2.735: TEPAT {TERLAMBAT=0, TEPAT=2}
|   |   |   |   |   |   IPS 7 ≤ 1.580: TERLAMBAT {TERLAMBAT=1, TEPAT=1}
|   |   |   |   IPS 6 ≤ 2.990: TEPAT {TERLAMBAT=0, TEPAT=51}
|   |   |   IPS 5 ≤ 2.635
|   |   |   IPS 3 > 2.480
|   |   |   |   IPS 1 > 2.920: TEPAT {TERLAMBAT=0, TEPAT=5}
|   |   |   |   IPS 1 ≤ 2.920
|   |   |   |   |   IPS 3 > 3.075: TEPAT {TERLAMBAT=0, TEPAT=2}
|   |   |   |   |   IPS 3 ≤ 3.075: TERLAMBAT {TERLAMBAT=6, TEPAT=0}
|   |   |   IPS 3 ≤ 2.480: TEPAT {TERLAMBAT=0, TEPAT=11}
```



## 5. Evaluation

- Atribut atau faktor yang **paling berpengaruh** adalah Status Mahasiswa, IPS2, IPS5, IPS1
- Atribut atau faktor yang **tidak berpengaruh** adalah Nama, Jenis Kelamin, Umur, IPS6, IPS7, IPS8



## 6. Deployment

- Budi membuat **program peningkatan disiplin dan pendampingan ke mahasiswa di semester awal (1-2) dan semester 5**, karena faktor yang paling menentukan kelulusan mahasiswa ada di dua semester itu
- Budi membuat **peraturan melarang mahasiswa bekerja paruh waktu di semester awal perkuliahan**, karena beresiko tinggi di kelulusan tepat waktu
- Budi membuat **program kerja paruh waktu di dalam kampus**, sehingga banyak pekerjaan kampus yang bisa intens ditangani, sambil mendidik mahasiswa supaya memiliki pengalaman kerja. Dan yang paling penting mahasiswa tidak meninggalkan kuliah karena pekerjaan
- Budi **memasukkan pola dan model yang terbentuk ke dalam sistem informasi akademik**, dimana sistem dibuat cerdas, sehingga bisa mengirimkan email analisis pola secara otomatis ke mahasiswa sesuai profilnya

# Course Outline

## 1. Pengantar Penelitian

- 1.1 Definisi Penelitian
- 1.2 Klasifikasi Penelitian
- 1.3 Gaya Penelitian Computing
- 1.4 Kontribusi dan Orisinalitas

## 2. Tahapan Penelitian

- 2.1 Tahapan Penelitian Umum
- 2.2 Tahapan Penelitian Computing
- 2.3 Tahapan Penelitian Computing  
Fokus Perbaikan Algoritma

## 3. Literature Review

- 3.1 Literatur Ilmiah
- 3.2 Teknik Mengelola Paper
- 3.3 Teknik Mereview Paper



# 1. Pengantar Penelitian

1.1 Definisi Penelitian

1.2 Klasifikasi Penelitian

1.3 Gaya Penelitian Bidang Computing

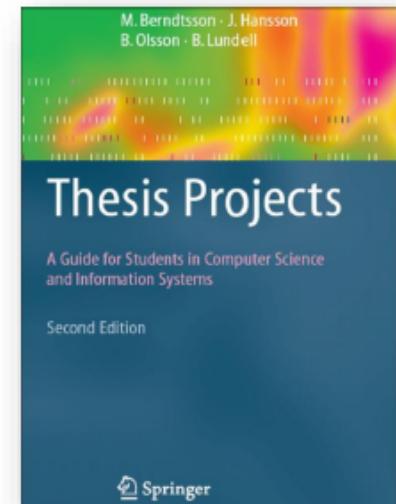
1.4 Kontribusi dan Originalitas



## 1.1 Definisi Penelitian

# Mengapa Melakukan Penelitian?

- Berangkat dari adanya **masalah penelitian**
  - yang mungkin sudah diketahui metode pemecahannya
  - tapi belum diketahui **metode pemecahan yang lebih baik**
- Research (Inggris) dan recherche (Prancis)
  - **re** (kembali)
  - **to search** (mencari)
- The process of exploring the unknown, studying and learning new things, **building new knowledge** about things that **no one has understood before**  
*(Berndtsson et al., 2008)*



# Apa Yang Dikejar di Penelitian?

Research is a **considered** activity, which aims to make an **original contribution** to knowledge

*(contribution to the body of knowledge, in the research field of interest)*

# Bentuk Kontribusi ke Pengetahuan

Kegiatan penyelidikan dan investigasi terhadap suatu masalah yang dilakukan secara berulang-ulang dan sistematis, dengan tujuan untuk menemukan atau merevisi teori, metode, fakta, dan aplikasi

(Berndtsson et al., 2008)



## Kontribusi ke Pengetahuan vs Kontribusi ke Masyarakat

**Kontribusi ke masyarakat** tidak secara langsung bisa diukur, karena itu tidak dimasukkan ke tujuan penelitian, tapi ke **manfaat penelitian**

# Apa itu Penelitian?

- Penelitian dilakukan karena ada **masalah penelitian**, dimana masalah penelitian sendiri muncul karena adanya **latar belakang masalah**, yang terlahir dari **masalah kehidupan**
- Penelitian dilakukan secara **terencana, sistematis, berulang-ulang** dan **terukur**
- Penelitian harus memiliki **orisinalitas (originality)** dan **kebaruan (novelty)**, serta menghasilkan **kontribusi yang orisinil** pada **pengetahuan** dalam bentuk **menemukan atau merevisi teori, metode, fakta, dan aplikasi**

# Pengembangan Software vs Penelitian

- Membangun software **bukanlah tujuan utama penelitian**, hanya *testbed* untuk mempermudah kita dalam mengukur hasil penelitian
  - Tidak ada *listing code*, UML atau screenshot software di paper-paper journal (SCOPUS/WoS), kecuali penelitian tentang perbaikan paradigma pemrograman, analisis design, dsb
- Ketika pada penelitian kita **mengusulkan perbaikan suatu algoritma (proposed method)**
  - Bidang image processing, topik penelitian face recognition, memikirkan *perbaikan metode/algoritma untuk pengenalan wajah* dengan akurat/efisien
  - Bidang data mining, topik decision tree, memikirkan *perbaikan algoritma decision tree* sehingga bisa memprediksi (klasifikasi) dengan lebih akurat
  - Untuk mempermudah eksperimen dan evaluasi, kita *menulis kode program (software)* untuk menguji dan mengevaluasi performance dari algoritma yang kita usulkan



# Karakter Peneliti

- Peneliti itu **boleh salah**
  - salah hipotesis
  - salah analisis
  - salah pengujian hipotesis
  - dsb
- Tapi peneliti **tidak boleh bohong atau menipu**
  - mempermainkan data
  - manipulasi hasil pengolahan statistik
  - dsb



## 1.2 Klasifikasi Penelitian

# Klasifikasi Penelitian

## 1. Pendekatan

1. Pendekatan **Kualitatif**
2. Pendekatan **Kuantitatif**

## 2. Metode

1. Metode Penelitian **Tindakan**
2. Metode **Eksperimen**
3. Metode **Studi Kasus**
4. Metode **Survei**

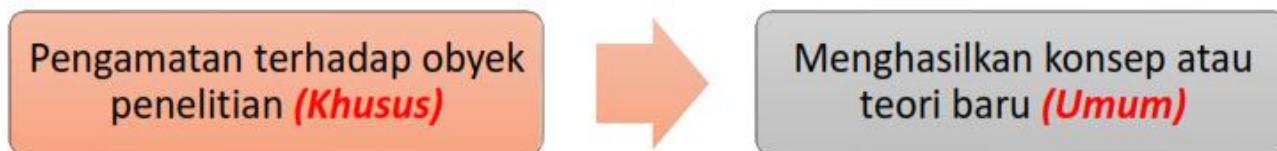
## 3. Jenis

1. Dasar vs Terapan
2. Eksplanatori vs Konfirmatori
3. Deskripsi vs Eksperimen vs Korelasi

# 1. Pendekatan

## 1. Pendekatan Kualitatif:

- Dari ilmu sosial, konsepnya **peningkatan pemahaman terhadap sesuatu**, dan bukan membangun penjelasan dari sesuatu (*Berndtsson et al., 2008*)
- Sifatnya **subyektif**, berorientasi ke observasi tanpa dikontrol, dan secara umum **tidak ada generalisasi** (*Blaxter, Hughes, & Tight, 2006*)
- Dilakukan **bidang sistem informasi**, dengan metode penelitian seperti “studi kasus” dan “survei”, berbasis pola alur **induktif**



## 2. Pendekatan Kuantitatif:

- Dari ilmu alam, konsepnya bagaimana sesuatu dibangun dan bekerja, dan **membangun penjelasan dari sesuatu**
- Sifatnya **obyektif**, berorientasi ke verifikasi, observasi yang dikontrol, dan secara umum **ada generalisasi** (*Blaxter et al., 2006*)
- Menggunakan skala numerik, berbasis pola alur **deduktif-induktif**



(*Berndtsson et al., 2008*)

## 2. Metode

### 1. Penelitian Tindakan

- Studi berupa monitoring dan pencatatan penerapan sesuatu oleh peneliti secara hati-hati, yang tujuannya untuk memecahkan masalah dan mengubah situasi (*Herbert, 1990*)
- Penelitian Tindakan Kelas (PTK) di bidang Pendidikan

### 2. Eksperimen

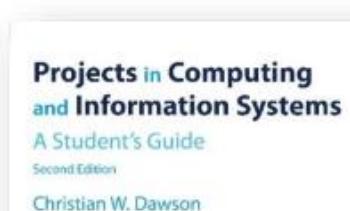
- Investigasi hubungan sebab akibat dengan menggunakan ujicoba yang dikontrol oleh peneliti
- Melibatkan pengembangan dan evaluasi
- Penelitian bidang Science dan Teknik

### 3. Studi Kasus

- Eksplorasi satu situasi secara mendalam dan hati hati (*Cornford and Smithson, 2006*)
- Penelitian bidang Sosial, Ekonomi, Politik

### 4. Survei

- Pengumpulan data dari populasi yang bisa diukur, dengan cara yang ekonomis (*Saunders et al., 2007*)
- Melibatkan penggunaan kuesioner dan interview



(Dawson, 2009)

### 3. Jenis

Deskripsi

Eksperimen

Korelasi

Kualitatif

Kuantitatif

Eksplanatori

Konfirmatori

Terapan

Dasar

# Penelitian Terapan



Penelitian Dasar

# Penerapan C4.5 untuk Prediksi Kelulusan Mahasiswa pada STMIK ABC



Teori Gain (*Kullback & Leibler, 1951*)

# Penerapan **Credal C4.5** untuk Prediksi Kelulusan Mahasiswa pada STMIK ABC

Split Criterion

**Credal C4.5**

**Imprecise  
Gain Ratio**

*(Mantas, 2013)*

**Imprecise Probability Theory (Walley, 1996)**



## 1.3 Gaya Penelitian Bidang Computing

# IEEE/ACM Computing Curricula 2005

## Computer Engineering (CE)

pengembangan sistem terintegrasi (software dan hardware)

Computer Engineer

## Information System (IS)

analisa kebutuhan dan proses bisnis serta desain sistem

System Analyst

## Information Technology (IT)

pengembangan dan pengelolaan infrastruktur IT

Infrastructure Engineer

## Computer Science (CS)

konsep computing dan pengembangan software

Computer Scientist

## Software Engineering (SE)

pengembangan software dan pengelolaan tahapan SDLC

Software Engineer

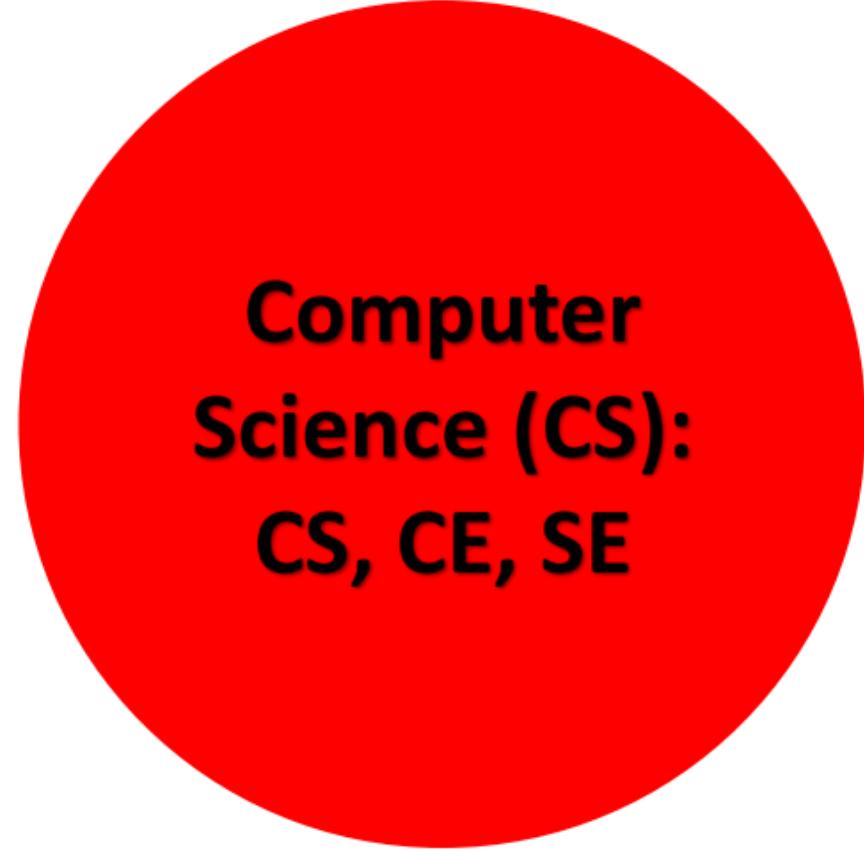


# Information Systems vs Computer Science



**Information  
Systems (IS):  
IS, IT**

Information systems specialists focus on **integrating information technology solutions and business processes** to meet the information needs of businesses and other enterprises  
*(ACM CC 2005)*



**Computer  
Science (CS):  
CS, CE, SE**

Computer science spans a wide range, from its **theoretical and algorithmic foundations** to cutting-edge developments in robotics, computer vision, intelligent systems, bioinformatics, and other exciting areas *(ACM CC 2005)*

# The Scope of Information Systems

Information Systems as a field of academic study encompasses the **concepts, principles, and processes for two broad areas** of activity within organizations:

1. Acquisition, deployment, management, and strategy **for information technology resources and services**  
→ *the information systems function; IS strategy, management, and acquisition; IT infrastructure; enterprise architecture; data and information*
2. Packaged system acquisition or system development, operation, and evolution of infrastructure and systems **for use in organizational processes**  
→ *project management, system acquisition, system development, system operation, and system maintenance*)

*(Curriculum Guidelines for Undergraduate Degree Programs in  
Information Systems, ACM and AIS, 2010)*



# Information Systems Profession

- IS specialists focus on **integrating information technology solutions and business processes** to meet the information needs of businesses and other enterprises, enabling them to achieve their objectives in an **effective, efficient way**
- IS specialists concerned with the information that computer systems can provide **to aid an enterprise in defining and achieving its goals**, and the processes that an enterprise can implement or improve using **information technology**
- They must **understand both technical and organizational factors**, and they must be able to help an organization determine how **information and technology-enabled business processes** can provide a competitive advantage

*(Computing Curricula 2005: The Overview Report, ACM and IEEE CS, 2006)*

# Computer Science Profession

1. They **design and implement software**. Computer scientists take on challenging programming jobs. They supervise other programmers, keeping them aware of new approaches (**Software Development**)
2. They **devise new ways to use computers**. Progress in the CS areas of **networking, database, and human-computer-interface** enabled the development of the World Wide Web. Now CS researchers are working with scientists from other fields to make robots become practical and intelligent aides, to use databases to **create new knowledge**, and to use computers to help **decipher the secrets of our DNA** (**Computing Algorithm Applications**)
3. They **develop effective ways to solve computing problems**. Computer scientists develop the best possible ways to store information in databases, send data over networks, and display complex images. Their theoretical background allows them to **determine the best performance possible**, and their study of algorithms helps them to develop new approaches that provide better performance (**Computing Algorithm Inventions**)

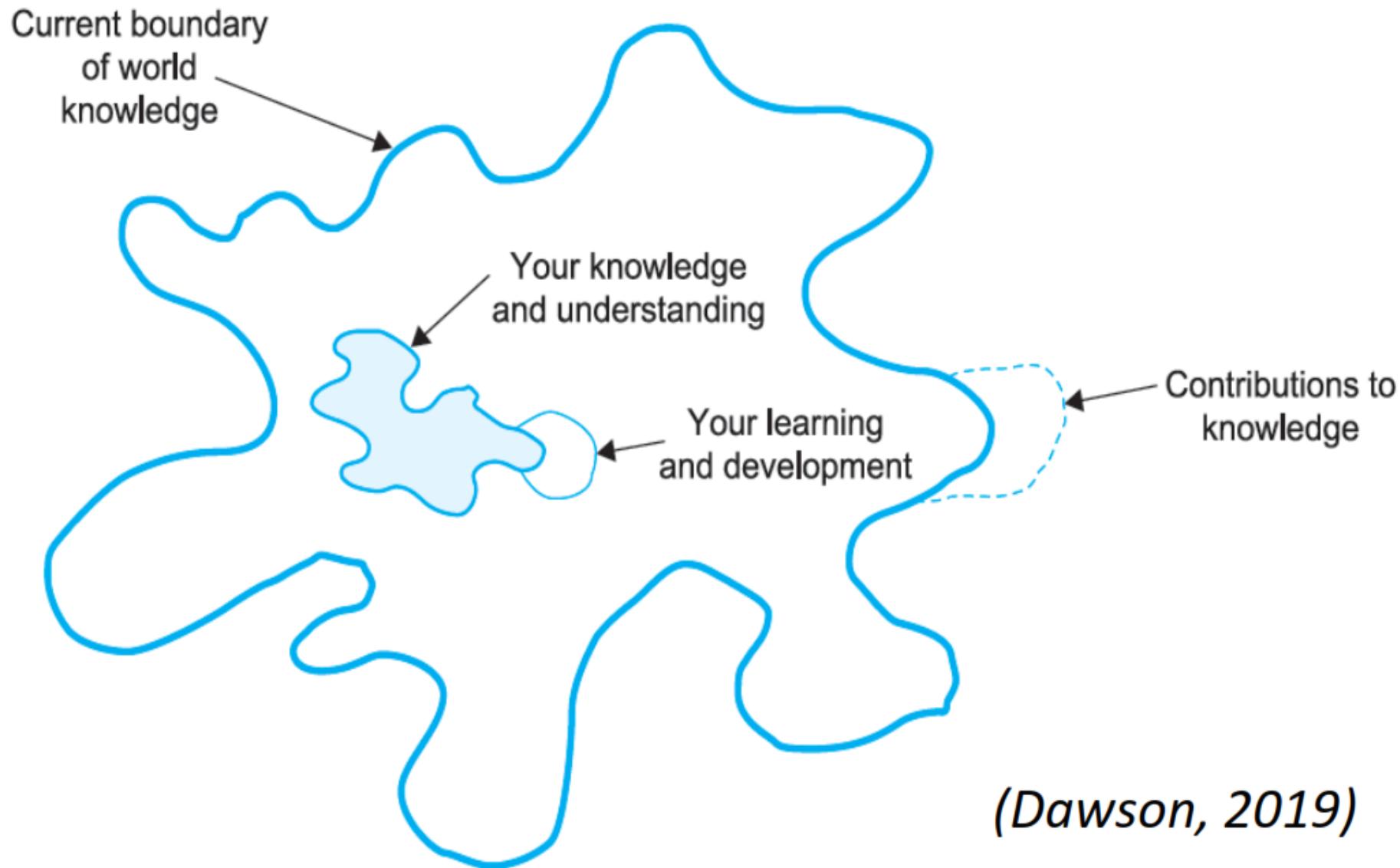


## 1.4 Kontribusi dan Orisinalitas

# Bentuk Kontribusi ke Pengetahuan

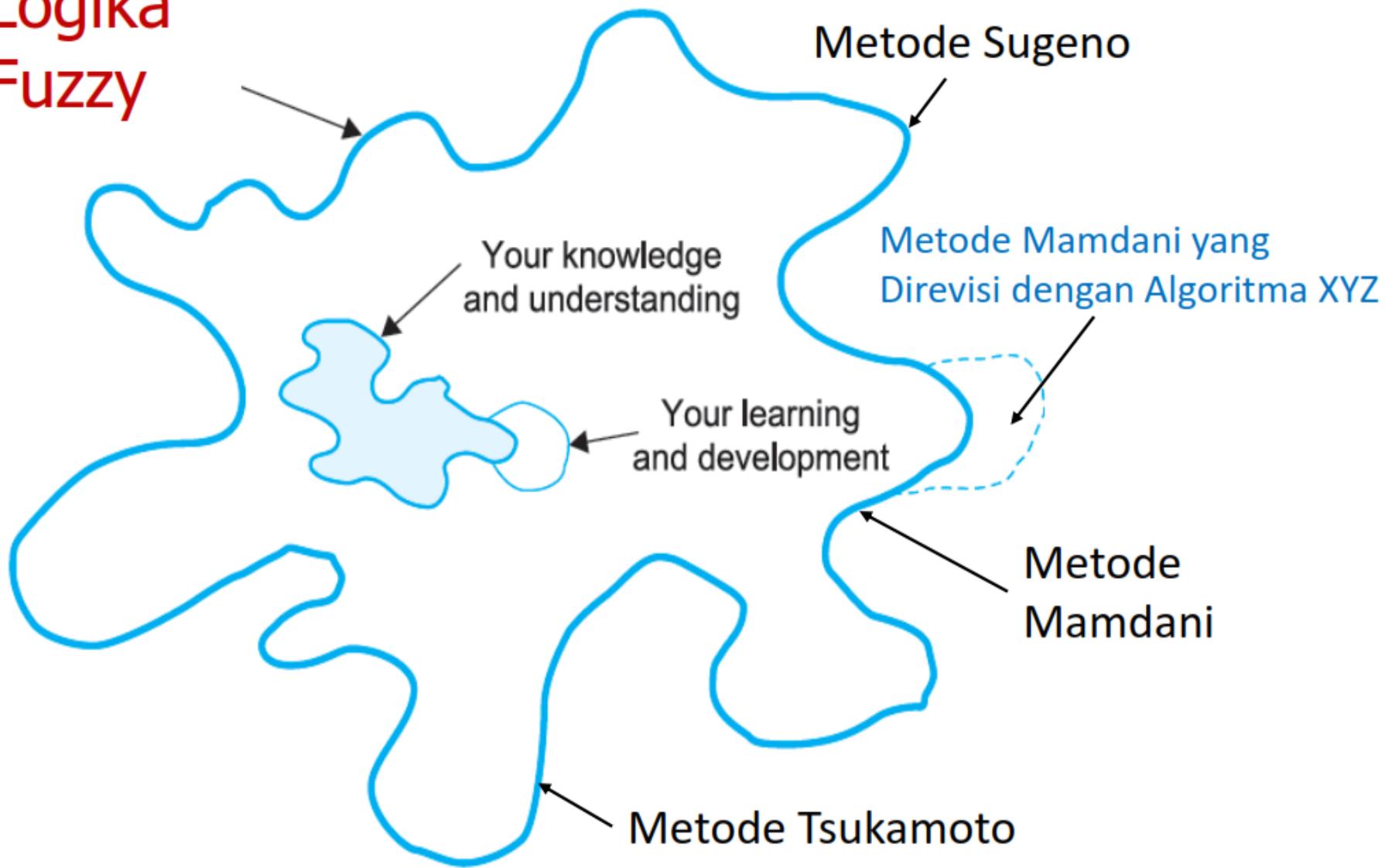
Kegiatan penyelidikan dan investigasi terhadap suatu masalah yang dilakukan secara berulang-ulang dan sistematis, dengan tujuan untuk menemukan atau merevisi teori, metode, fakta, dan aplikasi

# Bentuk Kontribusi ke Pengetahuan



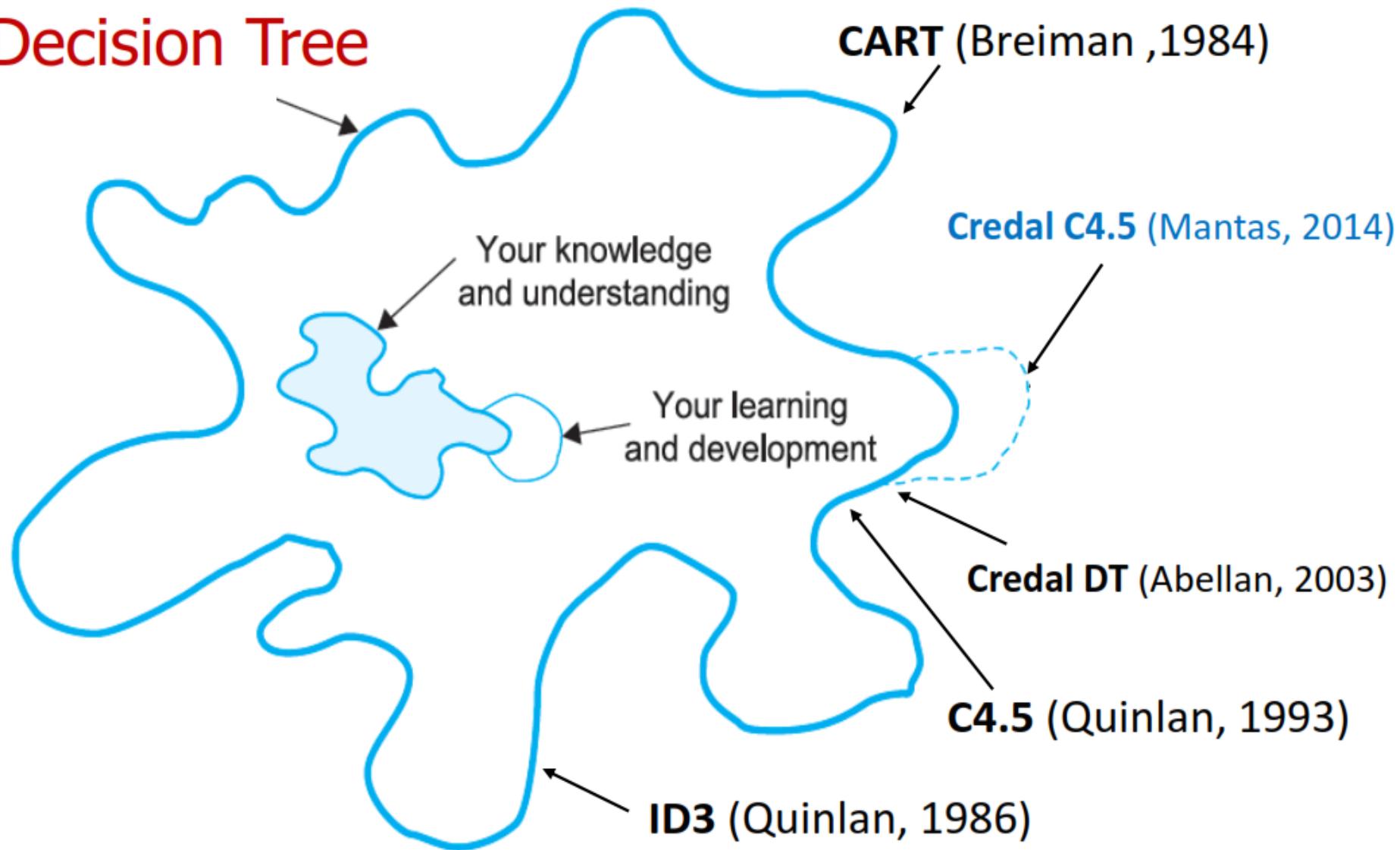
# Bentuk Kontribusi ke Pengetahuan

Logika  
Fuzzy



# Bentuk Kontribusi ke Pengetahuan

## Decision Tree



# Orisinalitas Penelitian

## 1. Orisinalitas pada Metode:

- Memecahkan masalah yang orang lain sudah pernah mengerjakan sebelumnya, tapi dengan metode yang berbeda
- Model penelitian yang kontribusi ada pada method improvement

## 2. Orisinalitas pada Masalah:

- Memecahkan suatu masalah yang orang lain belum pernah mengerjakan sebelumnya
- Model penelitian yang kontribusi ada pada penemuan masalah baru sebagai obyek penerapan metode

(Dawson, 2009)



## Contoh Tema Penelitian

**Algoritma Genetika untuk  
Penentuan Desain Bendungan  
yang Paling Optimal**

# Contoh Kontribusi pada Metode

- **Judul:**  
Penerapan Metode XYZ untuk Pemecahan Masalah Konvergensi Prematur pada Algoritma Genetika untuk Penentuan Desain Bendungan
- **Kontribusi:** Menerapkan Metode XYZ yang sebelumnya tidak pernah digunakan orang untuk memecahkan masalah konvergensi premature pada Algoritma Genetika

# Contoh Kontribusi pada Masalah

- **Judul:**  
Penerapan Algoritma Genetika untuk  
Penentuan Desain Bendungan dengan  
Tujuh Parameter
- **Kontribusi:** Penentuan Desain Bendungan  
dengan Tujuh Parameter (kebanyakan  
peneliti menggunakan tiga parameter)

# Contoh Kontribusi pada Masalah dan Metode

- **Judul:**

Penerapan **Metode XYZ** untuk Pemecahan Masalah Konvergensi Prematur pada Algoritma Genetika untuk Penentuan Desain Bendungan dengan **Tujuh Parameter**

- **Kontribusi:**

1. Penerapan **metode XYZ** untuk memecahkan masalah konvergensi premature pada algoritma genetika
2. Penentuan Desain Bendungan dengan **Tujuh Parameter**

# Contoh Penelitian Tanpa Kontribusi

- Penerapan Algoritma Genetika untuk Penentuan Desain Bendungan **di Bendungan Jatiluhur**
- Penerapan Algoritma Genetika untuk Penentuan Desain Bendungan **di Bendungan Gajah Mungkur**
- Penerapan Algoritma Genetika untuk Penentuan Desain Bendungan **di Bendungan Karang Kates**

\* banyak peneliti computing di Indonesia yang terjebak dengan **penelitian tanpa kontribusi** dan hanya mengganti obyek tempat, akhirnya ditolak ketika publikasi ke journal internasional terindeks

# Komparasi Penelitian D3/D4 vs S1 vs S2 vs S3

Aspek	Tugas Akhir (D3/D4)	Skripsi (D4/S1)	Tesis (S2)	Disertasi (S3)
Level Kontribusi	Penguasaan Kemampuan Teknis	Pengujian Teori	Pengembangan Teori	Penemuan Teori Baru
Bentuk Kontribusi	Implementasi dan pengembangan	Implementasi dan pengembangan	Perbaikan Secara Inkremental dan Terus Menerus	Substansial dan Invention
Target Publikasi	-	Domestic Conference	International Conference	International Journal

(Permendikbud No 3 tahun 2020 tentang SNPT)

# Komparasi Kontribusi Penelitian S1 vs S2 vs S3

- D3/D4:
  - Pengembangan Sistem Informasi Rumah Sakit untuk Rumah Sakit “Suka Sembuh”
  - Karakter: *menguasai skill teknis*
- S1:
  - Sistem Cerdas Berbasis **Neural Network** untuk Prediksi Harga Saham
  - Karakter: *menguji teori, ada software development*
- S2:
  - Penerapan **Algoritma Genetika** untuk **Pemilihan Arsitektur Jaringan Secara Otomatis** pada **Neural Network** untuk Prediksi Harga Saham
  - Karakter: *mengembangkan teori (*perbaikan metode*), ada kontribusi ke teori/metode meskipun specific obyek*
- S3:
  - Penerapan **Algoritma XYZ** untuk **Pemilihan Arsitektur Jaringan Secara Otomatis** pada **Neural Network**
  - Karakter: *menemukan teori (*invensi metode*), ada kontribusi ke teori/metode dengan generalisasi lebih luas*

# Komparasi Kontribusi Penelitian S1 vs S2 vs S3

- **S1:**

- Pengaruh **4P** Marketing Mix pada Peningkatan Penjualan Perusahaan XYZ
- *Kontribusi: menguji dan menerapkan teori/hukum/model/metode*

- **S2:**

- Pengaruh **4P+3C** Marketing Mix pada Peningkatan Penjualan Perusahaan XYZ
- *Kontribusi: mengembangkan dan memperbaiki teori/hukum/model/metode*

- **S3:**

- Pengaruh **ABCD** Marketing Mix pada Peningkatan Penjualan Perusahaan
- *Kontribusi: mengembangkan dan menemukan (invention) teori/hukum/model/metode baru yang sifatnya lebih general*

# Penelitian Yang Memiliki Kontribusi?

- Penerapan algoritma genetika untuk penjadwalan mata kuliah 
- Penerapan algoritma genetika berbasis *guided local search strategies* untuk penjadwalan mata kuliah (Yang, 2011) 
- Penerapan algoritma C4.5 untuk penentuan kelulusan mahasiswa tepat waktu: *Studi Kasus STMIK XYZ* 
- Penerapan algoritma C4.5 dengan penghitungan entropi berbasis metode ABC untuk penentuan kelulusan mahasiswa tepat waktu 

*Hanya penelitian dengan kontribusi ke pengetahuan yang bisa menembus jurnal-jurnal internasional terindeks*

# Penelitian Yang Memiliki Kontribusi?

No	Judul	
1	Penerapan Neural Network untuk Prediksi Harga Saham pada Perusahaan ABC	
2	Pemilihan Arsitektur Jaringan pada Neural Network Secara Otomatis dengan Menggunakan Algoritma Semut	
3	Modifikasi Penghitungan Gain dan Entropi untuk Peningkatan Akurasi pada Algoritma C4.5	
4	Penerapan Framework TOGAF untuk Pengembangan Enterprise Architecture pada Organisasi ABC	
5	Penerapan Framework TOGAF yang Dimodifikasi untuk Pengembangan Enterprise Architecture pada Perusahaan Skala Kecil dan Menengah	
6	Penerapan COBIT untuk Tata Kelola Organisasi ABC	
7	Integrasi COBIT dan TOGAF untuk Tata Kelola Organisasi ABC yang Lebih Komprehensif	



## Penelitian yang Berkualitas Tinggi

Topik dan skalanya **kecil, fokus, dalam**, dan membawa pengaruh yang besar ke bidang penelitian kita



## 2. Tahapan Penelitian

2.1 Tahapan Penelitian Umum

2.2 Tahapan Penelitian Computing

2.3 Tahapan Penelitian Computing Fokus Perbaikan Algoritma



## 2.1 Tahapan Penelitian Umum



# Tahapan Penelitian Umum

1. Identifikasi Masalah
2. Perumusan Hipotesis
3. Pengujian Hipotesis dan Analisis
4. Kesimpulan

# Tahapan Penelitian Umum vs Tesis

Tahapan Penelitian	Susunan Tesis
1. Identifikasi <b>Masalah</b>	1. Pendahuluan: <ul style="list-style-type: none"><li>- Latar Belakang</li><li>- Rumusan Masalah</li><li>- Tujuan Penelitian</li><li>- Manfaat Penelitian</li></ul>
2. Perumusan <b>Hipotesis</b>	2. Landasan Teori: <ul style="list-style-type: none"><li>- Penelitian yang Berhubungan</li><li>- Landasan Teori</li><li>- Kerangka Pemikiran</li></ul>
3. <b>Pengujian Hipotesis</b> dan Analisis Hasil	3. Metodologi Penelitian: <ul style="list-style-type: none"><li>- Metode Penelitian</li><li>- Metode Pengumpulan Data</li><li>- Metode Analisis Data</li><li>- Metode Pengukuran Penelitian</li></ul>
4. <b>Kesimpulan</b>	4. Analisis Hasil dan Pembahasan  5. Kesimpulan dan Saran



## 2.2 Tahapan Penelitian Computing

# Tahapan Penelitian Computing

## Literature Review

1. Penentuan Bidang Penelitian (*Research Field*)



2. Penentuan Topik Penelitian (*Research Topic*)



3. Penentuan Masalah Penelitian (*Research Problem*)



4. Perangkuman Metode-Metode Yang Ada (*State-of-the-Art Methods*)



5. Penentuan Metode Yang Diusulkan (*Proposed Method*)



6. Evaluasi Metode Yang Diusulkan (*Evaluation*)



7. Penulisan Ilmiah dan Publikasi Hasil Penelitian (*Publications*)

\*<https://www.site.uottawa.ca/~bochmann/dsrg/how-to-do-good-research/>

\*<http://romisatriawahono.net/2013/01/23/tahapan-memulai-penelitian-untuk-mahasiswa-galau/>

# 1. Penentuan Bidang Penelitian

- Ingat kembali seluruh mata kuliah yang sudah kita terima di perkuliahan
- Bidang penelitian di disiplin computing:

Software Engineering	Data Mining
Image Processing	Computer Vision
Networking	Human Computer Interaction
Soft Computing	Information Retrieval
Bioinformatics	dsb

- Tentukan berdasarkan passion!
- Contoh: Saya memilih bidang Software Engineering (SE)

## 2. Penentuan Topik Penelitian

1. **Searching** di ScienceDirect.Com, Springerlink, IEEE Explore, Google (Scholar):
  - research **trends challenge topics** on NAMA BIDANG
2. Untuk mempercepat pembelajaran, temukan survey paper berbentuk **Tertiary Study** (SLR dari SLR), karena isinya sudah merangkumkan **satu bidang penelitian**
3. Lanjutkan penentuan topik penelitian dengan **menemukan suvey/review paper (SLR, SMS)**, karena survey/review paper yang masuk jurnal terindeks pasti **membahas satu topik penelitian**

# 1. Cari Tertiary Study di Bidang Software Engineering

Adva

Search tip

40 result

Set search

Refine by

Years

2020 (1)

2019 (5)

2018 (2)

Show more

Article type

Review article

Publication

Information (29)

Advanced

Journal

Computer

Trends in

Review article

Identifying information needs in software engineering: An update

Apostolos Argyropoulos

Review article

Guidelines for conducting systematic mapping studies in software engineering: An update

Information and Software Technology, Volume 55, Issue 12, December 2013, Pages 2049-2075

Kai Petersen, Sairam Vakkalaru

Review article

A systematic review of systematic review process research in software engineering

Information and Software Technology, Volume 55, Issue 12, December 2013, Pages 2049-2075

Barbara Kitchenham, Pearl Brereton

Review article

A systematic mapping study of service-oriented software engineering

Information and Software Technology, Volume 55, Issue 11, November 2013, Pages 1845-1859

Roberto E. Lopez-Herrejon, Lorraine M. Gómez, Barbara Kitchenham

Combining service-orientation and software product line engineering: A systematic mapping study

Information and Software Technology, Volume 55, Issue 11, November 2013, Pages 1845-1859

Bardia Mohabbati, Mohsen Asadi, Dragan Gašević, Marek Hatala, Hausi A. Müller

Review article

Social computing for software engineering

Computer Science Review, Volume 6, Issue 1, March 2012, Pages 1-22

Amalia Ardini, Mahmood Hosseini, Barbara Kitchenham

Tools used in Global Software Engineering: A systematic mapping review

Information and Software Technology, Volume 54, Issue 7, July 2012, Pages 663-685

Javier Portillo-Rodríguez, Aurora Vizcaíno, Mario Piattini, Sarah Beecham

Review article

Gamification in software engineering

Information and Software Technology, Volume 53, Issue 5, May 2011, Pages 440-455

Oscar Pedreira, Félix García, Tore Dybå

Research synthesis in software engineering: A tertiary study

Information and Software Technology, Volume 53, Issue 5, May 2011, Pages 440-455

Daniela S. Cruzes, Tore Dybå

Software test-code engineering

Information and Software Technology, Volume 52, Issue 8, August 2010, Pages 806-820

Vahid Garousi Yusifoglu, Yaser Alavi, Barbara Kitchenham

Requirements engineering for software product lines: A systematic literature review

Information and Software Technology, Volume 52, Issue 8, August 2010, Pages 806-820

Vander Alves, Nan Niu, Carina Alves, George Valençap

Potential and limitations of systematic literature reviews in software engineering

Information and Software Technology, Volume 52, Issue 8, August 2010, Pages 792-805

Marta Fernández-Diego, Fernanda Gómez, Barbara Kitchenham, Rialette Pretorius, David Budgen, O. Pearl Brereton, ... Stephen Linkman

Systematic literature reviews in software engineering – A tertiary study

Information and Software Technology, Volume 52, Issue 8, August 2010, Pages 792-805

Barbara Kitchenham, Rialette Pretorius, David Budgen, O. Pearl Brereton, ... Stephen Linkman

## 2. Cari SLR dari Topik Penelitian yang Dipilih

- Setelah kita paham beberapa topik penelitian di bidang software engineering dari *Tertiary Study (SLR dari SLR)*
- Langkah berikutnya, kita **kumpulkan seluruh SLR** dengan keyword topik seperti di paper *Tertiary Study (SLR dari SLR)*
- Lanjutkan dengan **mengejar seluruh SLR dari topik yang kita akan angkat** pada penelitian kita

# Kiat Memilih Topik Penelitian

- Pilih topik **bukan karena pekerjaan kita sekarang**, tapi karena topiknya menarik (ada passion) dan secara penelitian dapat kita lakukan (tidak mission impossible)
- Usahakan cari penelitian yang membuat kita bisa **konsentrasi penuh ke method improvement**, tidak harus pontang-panting menjelaskan tentang obyek organisasi, mencari dataset, dsb
- Pilih topik yang **dataset sudah tersedia secara public**, jadi tidak perlu kita repot mencari dataset untuk eksperimen kita
- Pilih topik yang **mudah secara pengukuran penelitian** dan bila memungkinkan **pengukuran cukup dengan komputer**
  - Penelitian requirement engineering, termasuk yang **rumit pengukuran penelitiannya**, melibatkan manusia dan organisasi sebagai obyek
- Pilih topik **sesuai kapasitas dan kapabilitas**
  - Kita tidak mungkin penelitian tentang software process improvement apabila **tidak tersedia organisasi sebagai testbed** yang menerapkan metodologi yang kita kembangkan
- Pilih topik yang **memungkinkan kita lakukan dengan laptop** kita yang kita miliki sekarang, kecuali kita mendapatkan grant research besar yang memungkinkan pembelian infrastruktur penelitian
  - Penelitian global software engineering, software outsourcing, product line, relative agak perlu biaya lebih besar dan kompleks



## 3. Literature Review

3.1 Literatur Ilmiah

3.2 Teknik Mengelola Paper

3.3 Teknik Mereview Paper



### 3.1 Literatur Ilmiah

# Tahapan Penelitian Computing

## Literature Review

1. Penentuan Bidang Penelitian (*Research Field*)



2. Penentuan Topik Penelitian (*Research Topic*)



3. Penentuan Masalah Penelitian (*Research Problem*)



4. Perangkuman Metode-Metode Yang Ada (*State-of-the-Art Methods*)



5. Penentuan Metode Yang Diusulkan (*Proposed Method*)



6. Evaluasi Metode Yang Diusulkan (*Evaluation*)



7. Penulisan Ilmiah dan Publikasi Hasil Penelitian (*Publications*)

\*<https://www.site.uottawa.ca/~bochmann/dsrg/how-to-do-good-research/>

\*<http://romisatriawahono.net/2013/01/23/tahapan-memulai-penelitian-untuk-mahasiswa-galau/>

# Manfaat Mereview Literatur

- Memperdalam pengetahuan tentang bidang dan topik yang diteliti (*Textbooks*)
- Memperdalam pengetahuan tentang topik lebih detail yang diteliti (*Survey Paper*)
- Mengetahui hasil penelitian yang berhubungan dan yang sudah pernah dilaksanakan (Related Research) (*Technical Paper*)
- Mengetahui perkembangan ilmu pada bidang yang kita pilih (*state-of-the-art*) (*Technical atau Survey Paper*)
- Memperjelas masalah penelitian (*Technical Paper*)



# Jenis Literatur Ilmiah

- 1. Paper dari Journal \***
2. Paper dari Book Chapter
3. Paper dari Conference (Proceedings)
4. Thesis dan Disertasi
5. Report (Laporan) dari Organisasi yang Terpercaya
6. Buku Textbook

# Organisasi Yang Mengindeks Journal

1. **Thomson Reuters Web of Science (Clarivate Analytics)**
  - Since 1963, formerly produced by ISI, 12032 journals are indexed
  - Pengindeks journal yang memiliki level paling baik
  - <http://wokinfo.com>
2. **Scopus**
  - Launched by Elsevier in 2004, 20000 journals, conference papers and other are indexed
  - Pengindeks journal level standard, biasa untuk syarat menyelesaikan PhD
  - <http://scopus.com>
3. **Google Scholar**
  - Launched in 2004, mengindeks semua publikasi ilmiah yang online
  - <http://scholar.google.com>

# Algoritma Perangkingan Journal

## 1. Journal Impact Factor (JIF)

- Data source: ISI Web of Science

## 2. Eigenfactor Score (ES)

- Data source: ISI Web of Science
- <http://www.eigenfactor.org>

## 3. Scimago Journal Rank (SJR)

- Data source: Scopus
- <http://www.scimagojr.com>

## 4. Source Normalized Impact per Paper (SNIP)

- Data source: Scopus
- <http://www.journalindicators.com>

## 5. h-index

- Data source: Google Scholar
- <http://scholar.google.com/intl/en/scholar/metrics.html>

\* JIF adalah algoritma yang digunakan oleh ISI, sedangkan SJR adalah algoritma yang digunakan oleh SCOPUS

# Sumber Pencarian Literatur

## GRATIS

### Journal

- <http://scholar.google.com>
- <http://citeseer.ist.psu.edu>

## BERBAYAR

### Journal

- <http://sciencedirect.com>
- <http://www.ebscohost.com>
- <http://link.springer.com>
- <http://ieeexplore.ieee.org>
- <http://dl.acm.org>



Software defect prediction



## Cendekia

Sekitar 168.000 hasil (0,08 dtk)



### Artikel

Koleksiku

### Kapan saja

Sejak 2015

Sejak 2014

Sejak 2011

Rentang khusus...

### Urutkan menurut relevansi

Urutkan menurut tanggal

sertakan paten

mencakup kutipan

Buat lansiran

Kiat: Telusuri laman berbahasa **Bahasa Indonesia saja**. Anda dapat menentukan bahasa penelusuran di [Setelan Cendekia](#).

### A critique of software defect prediction models

[NE Fenton, M Neil](#) - [Software Engineering, IEEE Transactions ...](#), 1999 - [ieeexplore.ieee.org](#)

Abstract—Many organizations want to predict the number of defects (faults) in **software** systems, before they are deployed, to gauge the likely delivered quality and maintenance effort. To help in this numerous **software** metrics and statistical models have been ...

Dirujuk 887 kali Artikel terkait 12 versi Kutip Simpan

[PDF] dari [qmull.ac.uk](#)

### Benchmarking classification models for software defect prediction: A proposed framework and novel findings

[S Lessmann, B Baesens, C Mues...](#) - [Software Engineering, ....](#), 2008 - [ieeexplore.ieee.org](#)

Abstract—**Software defect prediction** strives to improve **software** quality and testing efficiency by constructing predictive classification models from code attributes to enable a timely identification of fault-prone modules. Several classification models have been evaluated ...

Dirujuk 352 kali Artikel terkait 12 versi Kutip Simpan

[PDF] dari [lums.edu.p...](#)

### Empirical assessment of machine learning based software defect prediction techniques

[VUB Challagulla, FB Bastani, IL Yen...](#) - [International Journal on ...](#), 2008 - [World Scientific](#)

Automated reliability assessment is essential for systems that entail dynamic adaptation based on runtime mission-specific requirements. One approach along this direction is to monitor and assess the system using machine learning-based **software defect prediction** ...

Dirujuk 122 kali Artikel terkait 10 versi Kutip Simpan

### Software defect prediction

[BE Hunt Jr, JJ Kirkpatrick, RA Kloss...](#) - [US Patent ...](#), 2014 - [freepatentsonline.com](#)

Abstract: A method of **software defect prediction** by a computer is provided. The method comprises identifying **software** test organizations scheduled to perform testing on an application or applications, where the scope of the **software** testing varies between ...



### 3.3 Teknik Mereview Paper

# Jenis Paper Ilmiah

## 1. Technical Paper

1. Paper yang isinya adalah hasil penelitian dan eksperimen yang dilakukan seorang peneliti
2. Penilaian kualitas technical paper dari kontribusi ke pengetahuan

## 2. Survey Paper

1. Paper yang isinya adalah review dan survey tentang topik/tema suatu penelitian, biasanya jumlah penelitian yang direview mencapai ratusan atau ribuan
2. Rujukan dan panduan penting bagi peneliti yang baru memulai penelitian untuk memahami suatu topic/tema penelitian secara komprehensif



### 3.3.1 Technical Paper

# Kiat Mereview Technical Paper

## 1. Pahami Masalah Penelitian

- Apakah penelitian hanya menyelesaikan masalah yang dibuat-buat?
- Apakah masalah penelitian dilandasi dan divalidasi?

## 2. Pahami Kontribusi

- Apakah peneliti hanya mengulang hal yang sudah ada?
- Apakah peneliti menyadari literatur lain yang berhubungan dengan penelitiannya?
- Apa yang baru dan orisinil di paper itu (metodologi, algoritma, evaluasi, validasi, tool, dsb.)?

## 3. Pahami Validitas Kontribusi

- Apakah teori atau model yang diusulkan sudah terbukti benar? Tidak adakah kesalahan pada pembuktian?
- Adakah faktor-faktor aneh pada proses eksperimen penelitian?
- Apakah benchmark yang dilakukan realistik atau hanya buatan? Ataukah membandingkan apel dan jeruk?
- Apakah generalisasi cukup valid?



# Masalah Penelitian

- Masalah penelitian adalah alasan utama mengapa penelitian harus dilakukan
- Reviewer jurnal internasional menjadikan “masalah penelitian” sebagai parameter utama proses review
- Masalah penelitian harus objective (tidak subjective), dan harus dibuktikan secara logis dan valid bahwa masalah itu benar-benar masalah
- Supaya logis dan valid, perlu dilakukan objektifikasi masalah, dengan cara melandasi masalah penelitian dengan literature terbaru
- Dimana munculnya di paper:
  - Abstract
  - Introduction



# Alur Terbentuknya Masalah Penelitian

- Penelitian dilakukan karena ada masalah penelitian
- Dimana masalah penelitian sendiri muncul karena ada latar belakang masalah penelitian
- Latar belakang masalah penelitian itu berangkatnya bisa dari masalah kehidupan (obyek penelitian)

# Contoh Alur Latar Belakang Masalah Penelitian:

*Penerapan XYZ untuk Masalah E pada SVM untuk Prediksi Nilai Tukar Uang*

- Nilai tukar uang adalah faktor penting pada perekonomian suatu negara. Nilai tukar uang perlu diprediksi supaya kebijakan perekonomian bisa diambil dengan lebih akurat dan efisien...
- Metode untuk prediksi nilai tukar yang saat ini digunakan adalah regresi linier, neural network dan support vector machine...
- Regresi linier memiliki kelebihan A dan kelemahan B...
- Neural network memiliki kelebihan C dan kelemahan D...
- Support vector machine memiliki kelebihan bisa mengatasi masalah B (pada regresi linier) dan D (pada neural network)... tapi memiliki kelemahan E
- Masalah penelitian pada penelitian di atas?
  - Kebijakan perekonomian negara?
  - Prediksi nilai tukar uang?
  - Metode apa yang sebaiknya dipakai untuk prediksi nilai tukar?
- Masalah:** Support vector machine memiliki kelebihan memecahkan masalah B dan D (argumentasi dipilih), tapi **memiliki kelemahan E**
- Tujuan:** Menerapkan **metode XYZ** untuk memecahkan masalah E pada support vector machine

# Contoh Alur Latar Belakang Masalah Penelitian:

Penerapan XYZ untuk E pada Fuzzy Logic untuk Pengaturan Lampu Lalu Lintas Dinamis

- Kemacetan lalu lintas di kota besar semakin meningkat
- Penyebab kemacetan adalah traffic light persimpangan jalan
- Traffic light yang ada adalah statis (tetap waktunya) sehingga tidak dapat menyelesaikan kondisi kepadatan kendaraan yang berbagaimana waktu
- Traffic light harus didesain dinamis sesuai perubahan berbagai parameter
- Metode untuk menentukan waktu yang tepat secara dinamis dapat menggunakan AHP, ANP, Fuzzy Logic,
- AHP memiliki kelebihan A dan kelemahan B...
- ANP memiliki kelebihan C dan kelemahan D...
- Fuzzy logic memiliki kelebihan bisa mengatasi masalah B (pada AHP) dan D (pada ANP)... tapi memiliki kelemahan E
- Masalah penelitian pada penelitian di atas?
  - Bagaimana mengatasi kemacetan lalu lintas?
  - Bagaimana mendesain traffic light?
  - Metode apa yang sebaiknya dipakai untuk penentuan traffic light secara dinamis?
- **Masalah:** Fuzzy logic memiliki kelebihan memecahkan masalah B dan D (argumentasi dipilih), tapi **memiliki kelemahan E**
- **Tujuan:** Menerapkan **metode XYZ** untuk memecahkan masalah E pada fuzzy logic

# Contoh Masalah Penelitian

- **Ungu:** Obyek Data (Opsional, Bisa Dataset Publik)
- **Oranye:** Topik (Obyek Metode yang Diperbaiki)
- **Merah:** Masalah Penelitian
- **Hijau:** Metode Perbaikan yang Diusulkan
- **Biru:** Pengukuran Penelitian (Tidak Harus Masuk Judul)

Penerapan **Particle Swarm Optimization** untuk **Pemilihan Parameter** Secara Otomatis pada **Support Vector Machine** untuk **Prediksi Produksi Padi**

Research Problem (RP)	Research Question (RQ)	Research Objective (RO)
<p>SVM dapat memecahkan masalah ‘over-fitting’, lambatnya konvergensi, dan sedikitnya data training, akan tetapi memiliki kelemahan pada sulitnya pemilihan parameter SVM yang sesuai yang mengakibatkan akurasi tidak stabil</p>	<p>Seberapa meningkat akurasi metode SVM apabila PSO diterapkan pada proses pemilihan parameter?</p>	<p>Menerapkan PSO untuk pemilihan parameter yang sesuai pada SVM (C, lambda dan epsilon), sehingga hasil prediksinya lebih akurat</p>

# Contoh Masalah Penelitian

- **Ungu:** Obyek Data (Opsional, Bisa Dataset Publik)
- **Oranye:** Topik (Obyek Metode yang Diperbaiki)
- **Merah:** Masalah Penelitian
- **Hijau:** Metode Perbaikan yang Diusulkan
- **Biru:** Pengukuran Penelitian (Tidak Harus Masuk Judul)

- Masalah Penelitian (*Research Problem*):
  - Neural network terbukti memiliki performa bagus untuk menangani data besar seperti pada data prediksi harga saham, akan tetapi memiliki kelemahan pada pemilihan arsitektur jaringannya yang harus dilakukan secara trial error, sehingga tidak efisien dan mengakibatkan hasil prediksi kurang akurat
- Rumusan Masalah (*Research Question*):
  - Bagaimana peningkatan akurasi dan efisiensi neural network apabila pada pemilihan arsitektur jaringan diotomatisasi menggunakan algoritma genetika?
- Tujuan Penelitian (*Research Objective*):
  - Menerapkan algoritma genetika untuk mengotomatisasi pemilihan arsitektur jaringan pada neural network sehingga lebih efisien dan hasil prediksi lebih akurat

# Contoh Masalah Penelitian

- **Ungu:** Obyek Data (Opsional, Bisa Dataset Publik)
- **Oranye:** Topik (Obyek Metode yang Diperbaiki)
- **Merah:** Masalah Penelitian
- **Hijau:** Metode Perbaikan yang Diusulkan
- **Biru:** Pengukuran Penelitian (Tidak Harus Masuk Judul)

- Research Problem (RP):

- Algoritma **K-Means** merupakan algoritma clustering yang populer karena efisien dalam komputasi, akan tetapi memiliki **kelemahan pada sulitnya penentuan K yang optimal** dan komputasi yang **tidak efisien** bila menangani data besar (Zhao, 2010)

- Research Question (RQ):

- Seberapa **efisien algoritma Bee Colony** bila digunakan untuk **menentukan nilai K yang optimal** pada **K-Means**?

- Research Objective (RO):

- Menerapkan **algoritma bee colony** untuk **menentukan nilai K yang optimal** pada **K-Means** sehingga **komputasi lebih efisien**

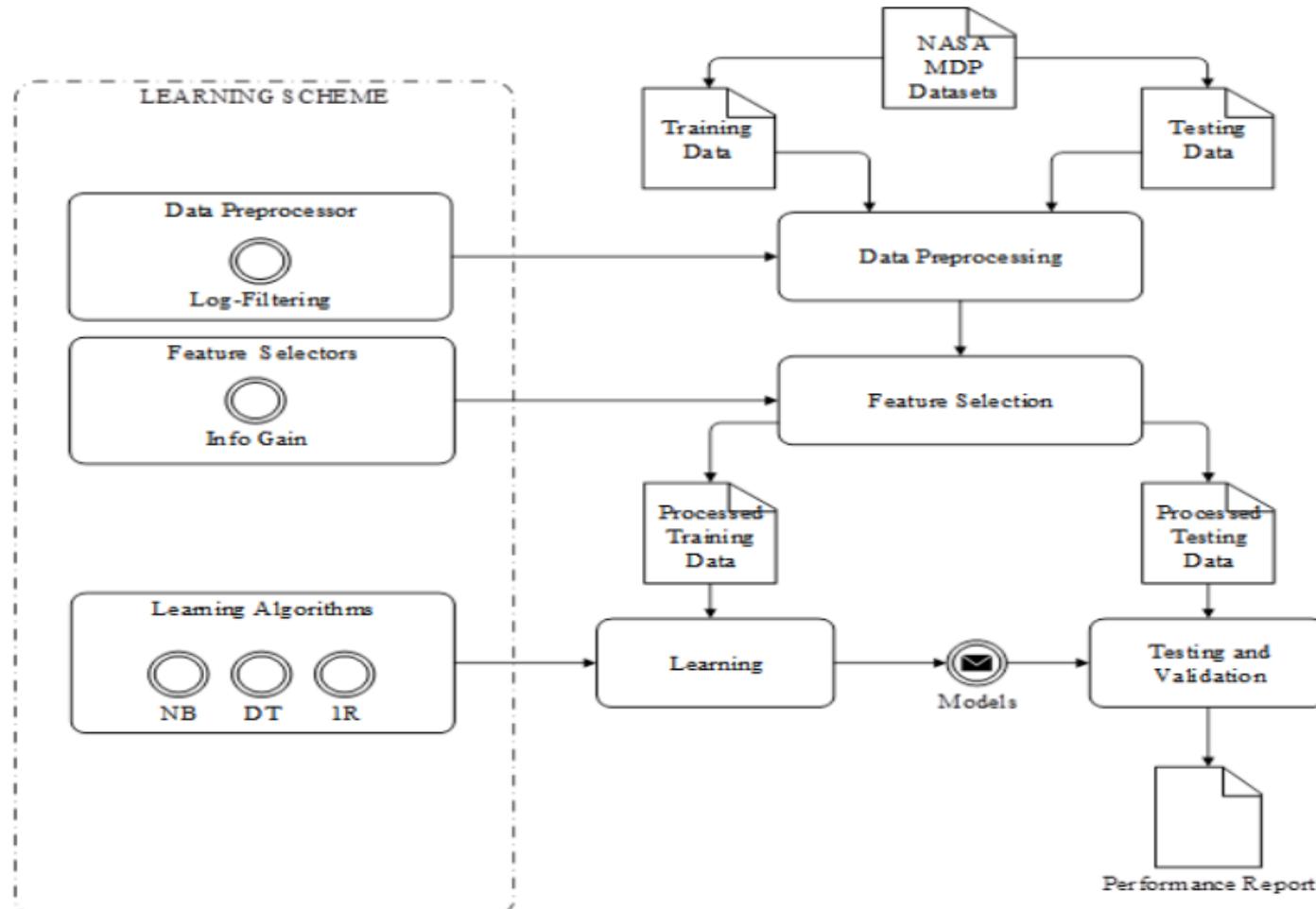
# Masalah Penelitian dan Landasannya

Masalah Penelitian	Landasan Literatur
Data set pada prediksi cacat software berdimensi tinggi, memiliki atribut yang bersifat noisy, dan classnya bersifat tidak seimbang, menyebabkan penurunan akurasi pada prediksi cacat software	There are <b>noisy data points</b> in the software defect data sets that can not be confidently assumed to be erroneous using such simple method ( <i>Gray, Bowes, Davey, &amp; Christianson, 2011</i> )
	The performances of software defect prediction improved when <b>irrelevant and redundant attributes</b> are removed ( <i>Wang, Khoshgoftaar, &amp; Napolitano, 2010</i> )
	The software defect prediction performance decreases significantly because the <b>dataset contains noisy attributes</b> ( <i>Kim, Zhang, Wu, &amp; Gong, 2011</i> )
	Software defect datasets have an <b>imbalanced nature</b> with very few defective modules compared to defect-free ones ( <i>Tosun, Bener, Turhan, &amp; Menzies, 2010</i> )
	<b>Imbalance</b> can lead to a model that is not practical in software defect prediction, because most instances will be predicted as non-defect prone ( <i>Khoshgoftaar, Van Hulse, &amp; Napolitano, 2011</i> )
	Software fault prediction data sets are often <b>highly imbalanced</b> ( <i>Zhang &amp; Zhang, 2007</i> )

# Formulasi RP-RQ-RO

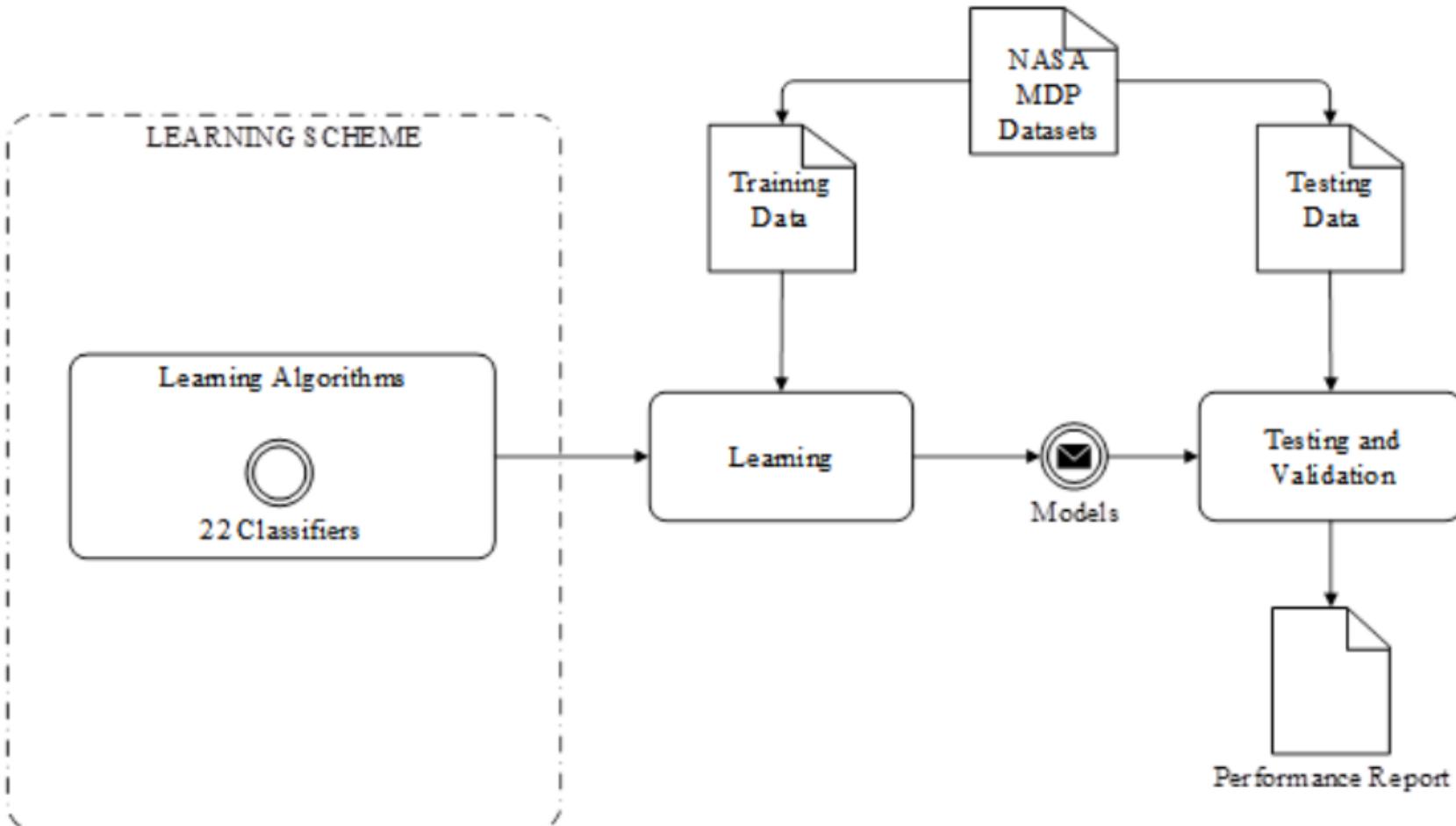
Research Problems (RP)		Research Questions (RQ)		Research Objectives (RO)	
RP	Data set pada prediksi cacat software berdimensi tinggi, dan memiliki <b>atribut yang bersifat noisy</b> , serta <b>classnya bersifat tidak balance</b>	RQ1	Algoritma pemilihan fitur apa yang performanya terbaik untuk meyelesaikan masalah atribut yang noisy pada prediksi cacat software?	RO1	Untuk mengidentifikasi algoritma pemilihan fitur apa yang memiliki performa terbaik apabila digunakan untuk menyelesaikan masalah atribut yang noisy pada prediksi cacat software
		RQ2	Algoritma meta learning apa yang performanya terbaik untuk menyelesaikan masalah class imbalance pada prediksi cacat software?	RO2	Untuk mengidentifikasi algoritma meta learning apa yang memiliki performa terbaik apabila digunakan untuk menyelesaikan masalah class imbalance pada prediksi cacat software
		RQ3	Bagaimana pengaruh penggabungan algoritma pemilihan fitur dan metode meta learning apabila digunakan untuk prediksi cacat software?	RO3	Untuk mengembangkan algoritma baru yang menggabungkan algoritma pemilihan fitur dan meta learning untuk prediksi cacat software

# Menzies Framework (Menzies et al. 2007)



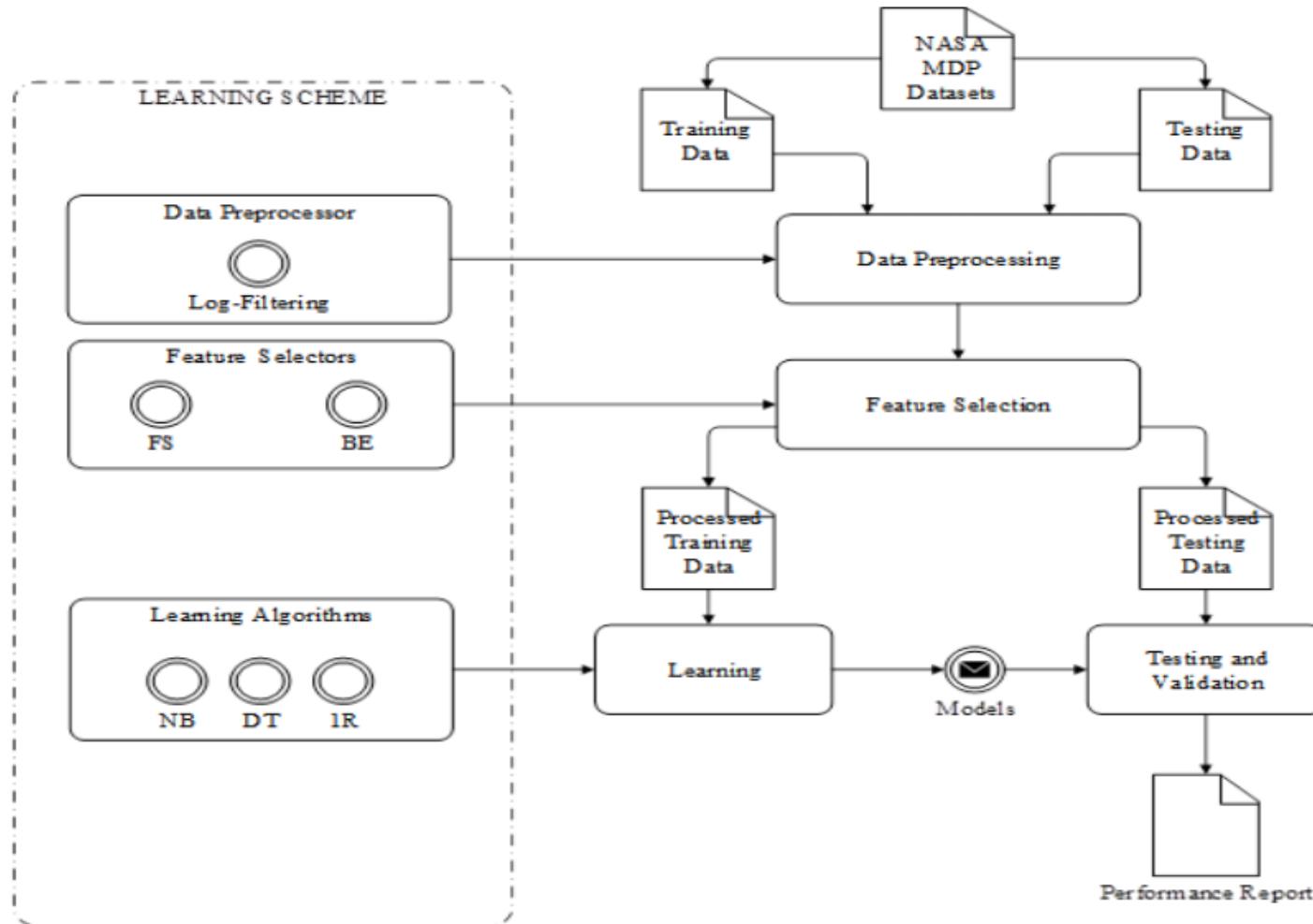
Framework	Dataset	Data Preprocessor	Feature Selectors	Meta-learning	Classifiers	Parameter Selectors	Validation Methods	Evaluation Methods
(Menzies et al. 2007)	NASA MDP	Log Filtering	Info Gain	-	267	3 algorithms (DT, 1R, NB)	-	10-Fold X Validation ROC Curve (AUC)

# Lessmann Framework (Lessmann et al. 2008)



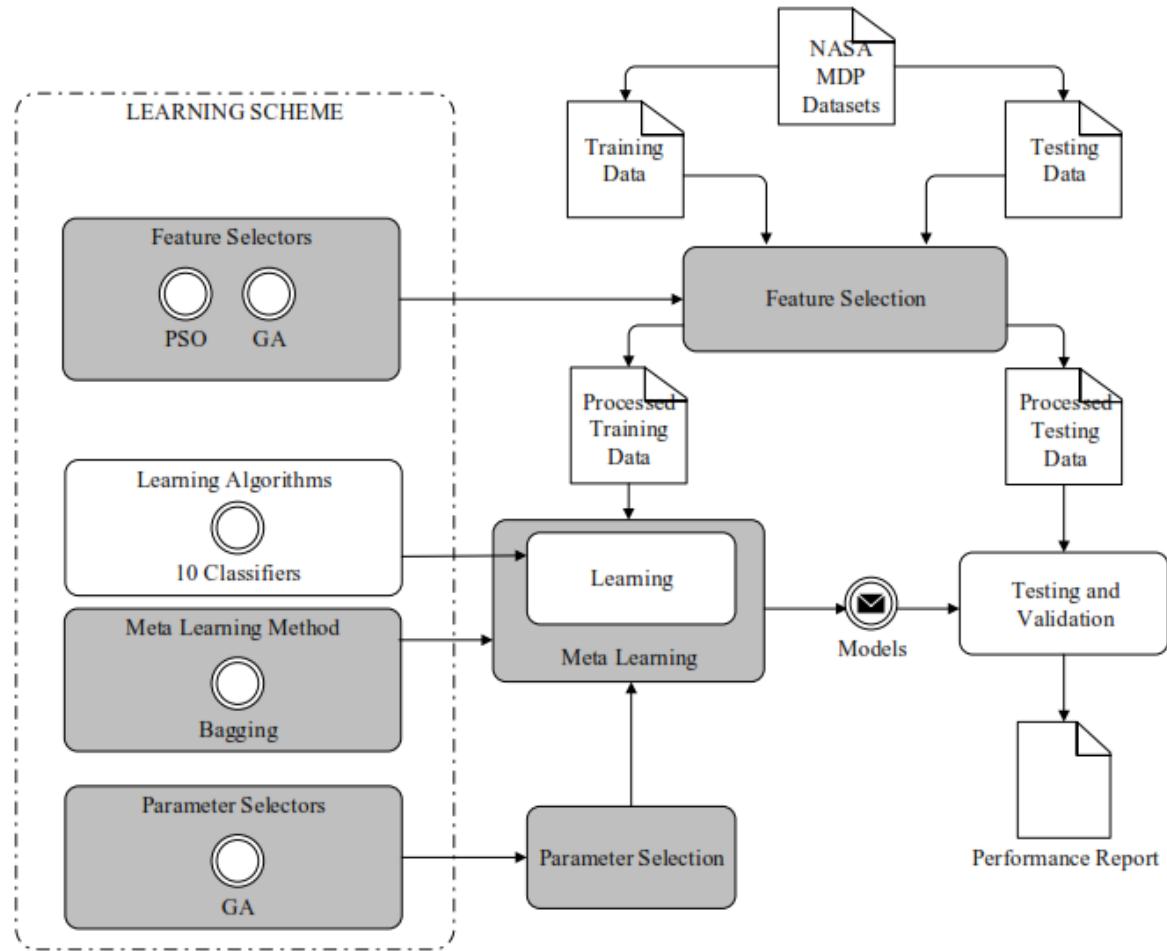
Framework	Dataset	Data Preprocessor	Feature Selectors	Meta-learning	Classifiers	Parameter Selectors	Validation Methods	Evaluation Methods
(Lessmann et al. 2008)	NASA MDP	-	-	-	268 algorithms	-	10-Fold X Validation	ROC Curve (AUC)

# Song Framework (Song et al. 2011)



Framework	Dataset	Data Preprocessor	Feature Selectors	Meta-learning	Classifiers	Parameter Selectors	Validation Methods	Evaluation Methods
(Song et al. 2011)	NASA MDP	Log Filtering	FS, BE	-	269	3 algorithms (DT, 1R, NB)	-	10-Fold X Validation ROC Curve (AUC)

# Proposed Framework

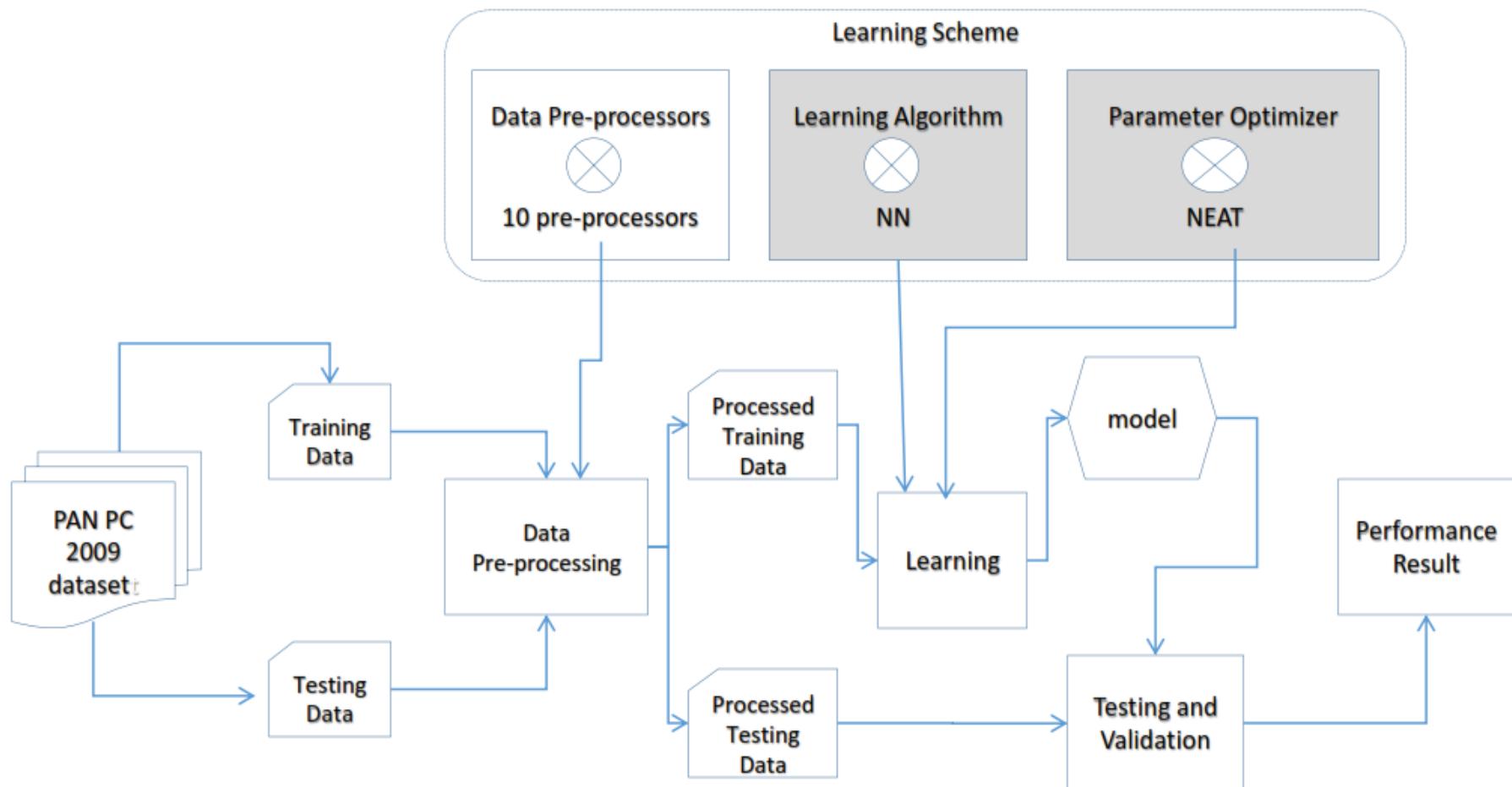


Framework	Dataset	Data Preprocessor	Feature Selectors	Meta-Learning	Classifiers	Parameter Selectors	Validation Methods	Evaluation Methods
(Menzies et al. 2007)	NASA MDP	Log Filtering	Info Gain		3 algorithm (DT, 1R, NB)	-	10-Fold X Validation	ROC Curve (AUC)
(Lessman et al. 2008)	NASA MDP	-	-		22 algorithm	-	10-Fold X Validation	ROC Curve (AUC)
(Song et al. 2011)	NASA MDP	Log Filtering	FS, BE		3 algorithm (DT, 1R, NB)	-	10-Fold X Validation	ROC Curve (AUC)
Proposed Framework	NASA MDP	-	PSO, GA	Bagging <sub>270</sub>	10 algorithms	GA	10-Fold X Validation	ROC Curve (AUC)

# Contoh State-of-the-Art Methods

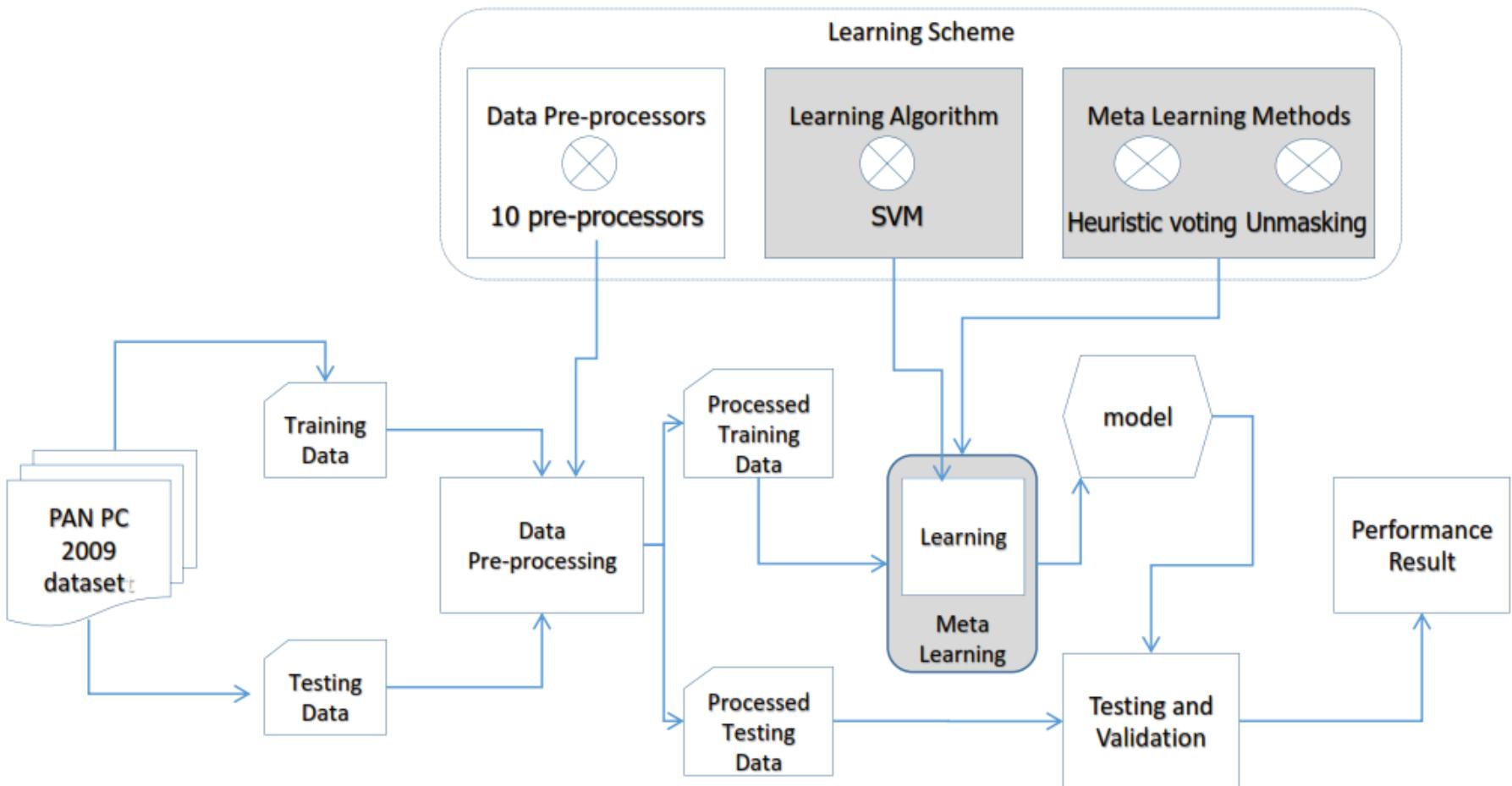
Model	Masalah Penelitian	Dataset	Pengukuran	Hasil
Jiang et al. (2007)	SPM memiliki tingginya penerimaan informasi warna kulit yang benar, akan tetapi tidak dapat menerima pada informasi fitur tekstur	Deteksi Kulit: Simulasi BrainWeb dan IBTD dataset	Akurasi segmentasi yang optimal (FPR and TPR)	false positive rate (FPR) = 6,2% true positive rate (TPR) = 92,97%
Wighton et al. (2011)	LDA tidak cukup akurat untuk mengidentifikasi dalam ukuran kernel yang berbeda	Deteksi Kulit: Simulasi BrainWeb dan IBTD dataset	Sensitivitas dan Spesifisitas (TN dan TP)	True Negative = 91% True Positive= 93%
Kawulok et al. (2013)	Performa dari SPM secara signifikan dapat mengalami "kebocoran" karena transisi halus antara kulit dan non-kulit	Deteksi Kulit: IBTD dataset	Analisa spasial yang baik ( $\delta_{fp}$ + $\delta_{fn}$ )	Detection Rate (DR) =94% false positive rate (FPR) = 34% false negative rate (FNR) = 6.13%
Muryan (2014)	SPM sulit untuk mendeteksi informasi fitur tekstur pada warna kulit dan kondisi pencahayaan  Algoritma LDA sulit mengekstrak pada kernel yang berbeda	Deteksi Kulit: IBTD dataset	Analisa spasial yang baik ( $\delta_{fp}$ + $\delta_{fn}$ )	?

# Curran's Model (Curran, 2010)



Model	Dataset	Data Pre-processor	Classifiers	Parameter Optimizer	Feature Creator	Meta Learning	Validation method	Evaluation Method
(Curran, 2010)	PAN PC 2009	10 Pre-processors	NN	NEAT 273	-	-	Fitness function	Accuracy

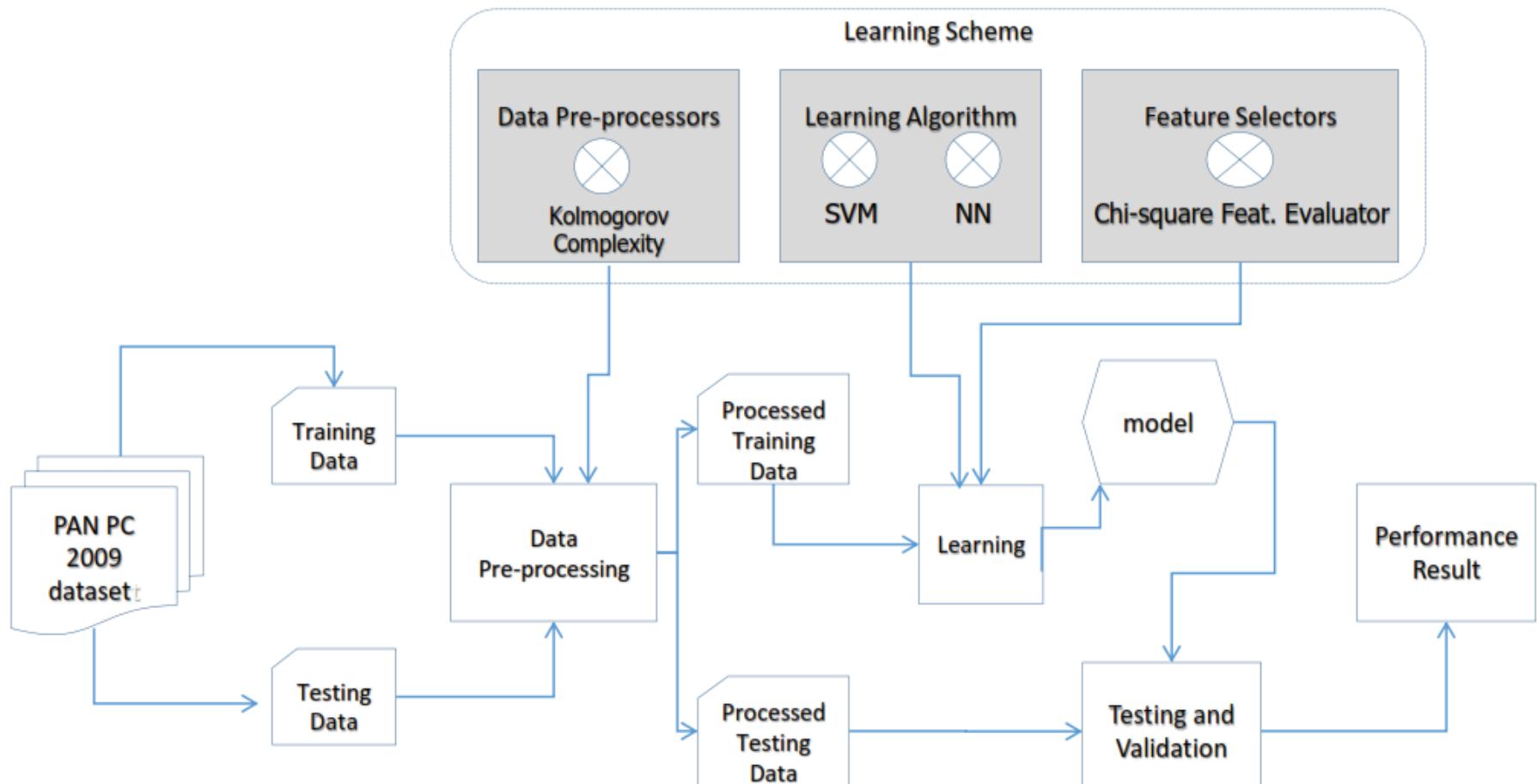
# Stein et al.'s Model (Stein et al., 2011)



Model	Dataset	Data Pre-processor	Classifiers	Parameter Optimizer	Feature Creator	Meta Learning	Validation method	Evaluation Method
(Stein et al., 2011)	PAN PC 2009	10 Pre-processors	SVM	- 274	-	Heuristic Voting, Unmasking	5-fold X validation	Recall Precision F-measure

# Seaward & Matwin's Model

(Seaward & Matwin, 2009)



Model	Dataset	Data Pre-processor	Classifiers	Parameter Optimizer	Feature Creator	Meta Learning	Validation method	Evaluation Method
(Seaward & Matwin, 2009)	PAN PC 2009	Kolmogorov Complexity	SVM, NN	- 275	FS: Chi-square Feat. Evaluator	-	-	Recall Precision F-measure

# Mulai Menulis Skripsi

# Bagian Skripsi

- **Judul**
- **Bab 1, Pendahuluan**
  - Latar Belakang
  - Rumusan Masalah
  - Tujuan Penelitian
  - Manfaat (Setelah dikerjaan skripsi ini apa manfaatnya?)
- **Bab 2, Tinjauan Pustaka**
  - *State of The Art* (pustaka dari 10 paper ilmiah)
  - Materi Khusus sesuai masalah
- **Bab 3, Perancangan / Perencanaan Sistem dll..**
  - Alur diagram Sistem / Algoritma / Model Data Mining
  - Desain Sistem
  - Contoh Perhitungan atau Penerapan Model
  - Contoh Evaluasi Model Data Mining
- Bab 4, Penerapan dan Analisis
- Bab 5, Kesimpulan

# Kiat Membuat Latar Belakang

- Link [Paper](#)
- **Kiat 1, PAHAMI DUA GAYA RESEARCH DI BIDANG COMPUTING**

Sebelumnya perlu dipahami bahwa gaya penelitian di bidang komputer (computing) secara umum terbagi dua yaitu gaya Computer Science (CS) dan gaya Information Systems (IS) (Berndtsson et al., 2008). CS memiliki karakteristik penelitian dan isu berhubungan dengan core technology dan perbaikan metode (method improvement). Sedangkan penelitian IS lebih cenderung ke arah isu tentang interaksi teknologi dan sosial, termasuk diantaranya mengukur dan menganalisa kesuksesan penerapan teknologi dan sistem informasi. Tulisan kali ini akan lebih cenderung ke alur latar belakang masalah penelitian bergaya CS, meskipun tetap bisa digunakan untuk penelitian IS.

- Link [Paper](#)
- Kiat 2, MENJAWAB SEMUA PERTANYAAN WHY DI JUDUL

Latar belakang masalah penelitian akan menjawab semua pertanyaan MENGAPA (*WHY*) dari judul penelitian kita. Untuk mempermudah penjelasan, saya akan gunakan, terjemahkan dan revisi paper penelitian (Fei et al, 2008) untuk contoh paper yang kita bahas. Karena judul penelitiannya adalah **Prediksi Produksi Padi dengan menggunakan Support Vector Machine berbasis Particle Swarm Optimization**, maka latar belakang masalah harus bisa menjawab pertanyaan:

1. mengapa padi?
2. mengapa prediksi produksi padi?
3. mengapa support vector machine?
4. mengapa particle swarm optimization?

Bagaimana cara menguraikan jawaban dari pertanyaan 1-4, akan disajikan dalam contoh latar belakang masalah di bawah.

- Link [Paper](#)
- Kiat 3, **POLA ALUR DAN POKOK PIKIRAN PARAGRAF**

Kunci dari keberhasilan menyusun latar belakang masalah penelitian seberapa komprehensif kita merangkumkan penelitian kita. Tulisan yang baik adalah bahwa dengan hanya membaca latar belakang masalah, orang langsung bisa memahami, apa yang kita lakukan pada penelitian kita. Untuk bisa mencapai itu, pokok pikiran seluruh paragraf pada latar belakang masalah penelitian harus memuat dan mengikuti 6 pola alur berikut. Untuk mempermudah mengingat, saya biasanya menggunakan singkatan **OMKKMASASOLTU**.

1. obyek penelitian (O)
  2. metode-metode yang ada (M)
  3. kelebihan dan kelemahan metode yang ada (KK)
  4. masalah pada metode yang dipilih (MASA)
  5. solusi perbaikan metode (SOL)
  6. rangkuman tujuan penelitian (TU)
- Contoh penerapan pola OMKKMASASOLTU ini, akan cepat dipahami melalui contoh latar belakang masalah yang saya uraikan di bawah

- **Prediksi Produksi Padi dengan menggunakan Support Vector Machine berbasis Particle Swarm Optimization**
- **Latar Belakang Masalah**
- *Padi adalah komoditas yang penting di china, karena tingkat produksinya tinggi (FAO Report, 2009) (1. mengapa padi?). Produksi padi perlu diprediksi dengan akurat, karena hasil prediksi yang akurat sangat penting untuk membuat kebijakan nasional (Traill, 2008) (2. mengapa prediksi produksi padi?). [1. obyek penelitian]*
- *Metode prediksi rentet waktu seperti Support Vector Machine (SVM) (Yongsheng, 2008), Neural Network (NN) (Tseng, 2007) dan Grey Model (GM) (Wu, 2007) diusulkan oleh banyak peneliti (Huifei, 2009) untuk prediksi produksi padi. [2. metode-metode yang ada]*
- *NN memiliki kelebihan pada prediksi nonlinear, kuat di parallel processing dan kemampuan untuk mentoleransi kesalahan, tapi memiliki kelemahan pada perlunya data training yang besar, over-fitting, lambatnya konvergensi, dan sifatnya yang local optimum (Rosario, 2007). GM punya kelebihan di tingginya akurasi prediksi meskipun menggunakan data yang sedikit, akan tetapi GM memiliki kelemahan pada prediksi data yang sifatnya naik turun secara fluktuatif seperti pada data produksi padi (Wu, 2007). [3. kelebihan dan kelemahan metode yang ada]*

- SVM dapat memecahkan masalah NN dan GM, yaitu over-fitting, lambatnya konvergensi, dan sedikitnya data training (Vapnik, 2005), yang mana ini tepat untuk karakteristik data produksi padi pada penelitian ini ([3. mengapa support vector machine?](#)). Tetapi SVM memiliki kelemahan pada sulitnya pemilihan parameter SVM yang optimal (Coussement, 2008). [[4. masalah pada metode yang dipilih](#)]
- Particle Swarm Optimization (PSO) adalah metode optimisasi yang terbukti efektif digunakan untuk memecahkan masalah optimisasi multidimensi dan multiparameter pada pembelajaran pada machine learning seperti di NN, SVM, dan classifier lain (Brits, 2009) ([4. mengapa particle swarm optimization?](#)). [[5. solusi perbaikan metode](#)]
- Pada penelitian ini PSO akan diterapkan untuk pemilihan parameter SVM yang sesuai dan optimal, sehingga hasil prediksi lebih akurat. [[6. rangkuman tujuan penelitian](#)]

# Kiat Rumusan Masalah Dan Tujuan Penelitian(1)

- Ketika kita telah berhasil menyusun latar belakang masalah yang baik, **masalah dan tujuan penelitian** sudah pasti dapat kita rangkumkan dengan baik.
- Di Indonesia masalah penelitian, biasanya dirangkumkan dalam format identifikasi masalah (*problem statement*) dan rumusan masalah (*research question*).
- Jadi sebagai kelanjutan dari latar belakang masalah di atas, kita bisa rangkumkan masalah dan tujuan penelitian sebagai berikut:

# Kiat Rumusan Masalah Dan Tujuan Penelitian(2)

- ***Masalah Penelitian (Problem Statement atau Research Problem):***
- *SVM adalah algoritma yang memiliki performa sangat baik untuk prediksi rentet waktu, karena dapat memecahkan masalah over-fitting, lambatnya konvergensi, dan sedikitnya data training. Tetapi SVM memiliki kelemahan pada sulitnya pemilihan parameter yang optimal karena harus dilakukan secara trial and error, sehingga menyebabkan tingkat akurasi prediksi menjadi rendah*

# Kiat Rumusan Masalah Dan Tujuan Penelitian(3)

- ***Pertanyaan Penelitian (Research Question):***
  1. *Seberapa tinggi akurasi metode SVM apabila PSO diterapkan pada proses pemilihan parameter yang optimal?*
  2. *Alternatif research question lain yang bisa digunakan adalah seperti di bawah:*
  3. *Bagaimana peningkatan akurasi SVM apabila PSO diterapkan pada proses pemilihan parameter yang optimal?*
  4. *Bagaimana pengaruh penerapan PSO pada pemilihan parameter yang optimal pada akurasi metode SVM?*

# Kiat Rumusan Masalah Dan Tujuan Penelitian(3)

- *Tujuan Penelitian (Research Objective):*
- *Menerapkan PSO untuk pemilihan parameter yang optimal pada SVM, sehingga dapat meningkatkan akurasi hasil prediksi*

# Progres Skripsi 1 (Pendahuluan)

- **Buat Latar Belakang**, Sesuai dengan Kiat 1-3  
Membuat Latar Belakang.
- **Buat Rumusan Masalah**.
- **Buat Tujuan Penelitian**.

# Progres Skripsi 2 (Tinjauan Pustaka)

- **Mencari paper ilmiah melalui google scholar** sebanyak 10 paper yang berkaitan dengan topik dan kasus skripsi yang diambil.
- **Review paper ilmiah**, 1 halaman review dan 1 halaman gambar skema. Untuk dijadikan Bab 2, Tinjauan Pustaka.
- Review materi khusus terkait sesuai dengan bidang/topik/masalah yang diangkat. Contoh, Masalah padi pada [Paper](#) (Metode Prediksi Padi) maka harus membahas mengenai Metode Prediksi.

# Progres Skripsi 3 (Perancangan / Perencanaan Sistem dll..)

- Alur diagram Sistem / Algoritma / Model Data Mining.
- Desain Sistem.
- Contoh Perhitungan atau Penerapan Model
- Contoh Evaluasi Model Data Mining

Materi Selesai, Semoga  
Skripsi Lancar. Amin..