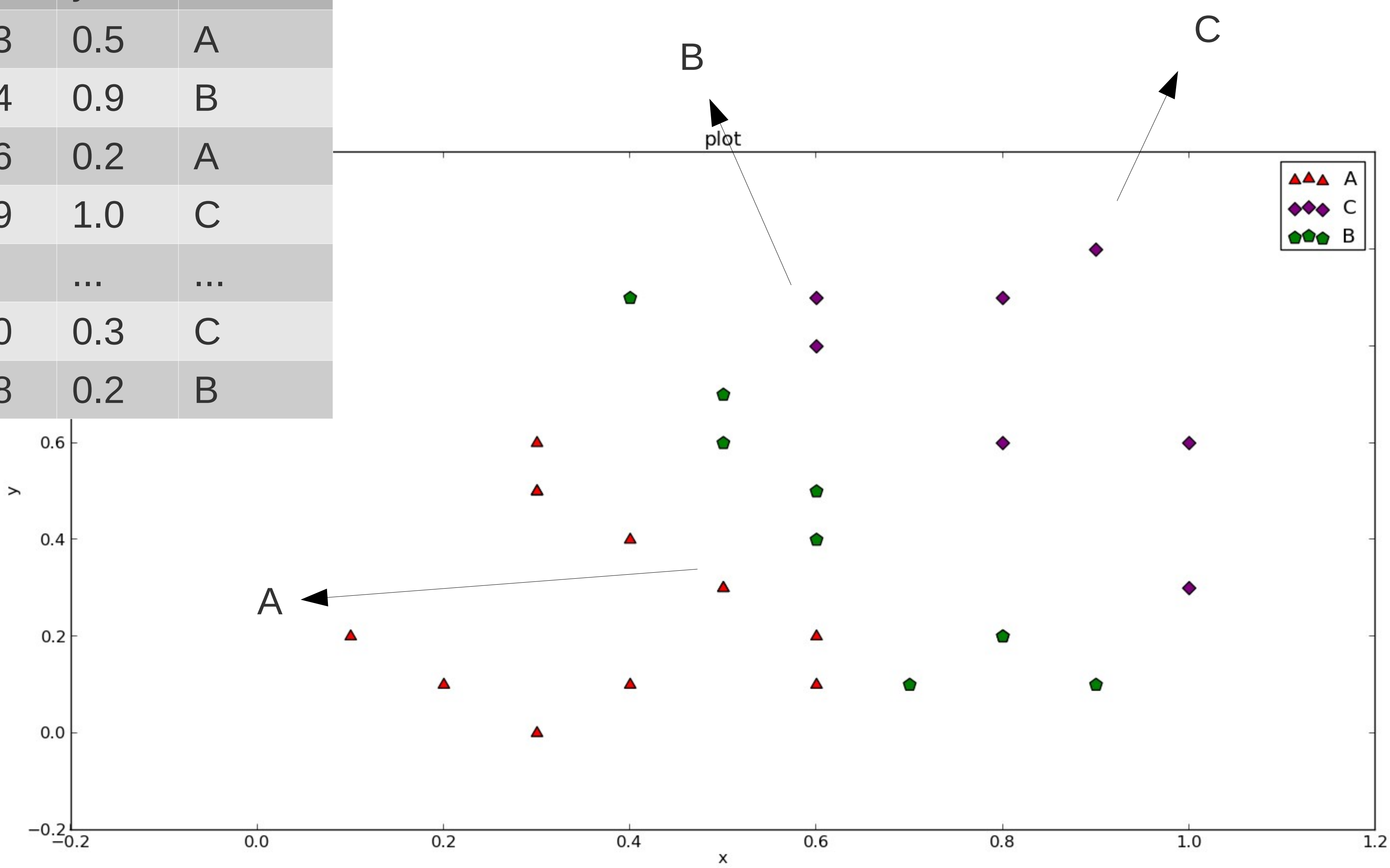
Grammatical Evolution untuk Ekstraksi Fitur dengan Pengukuran Multi Fitness

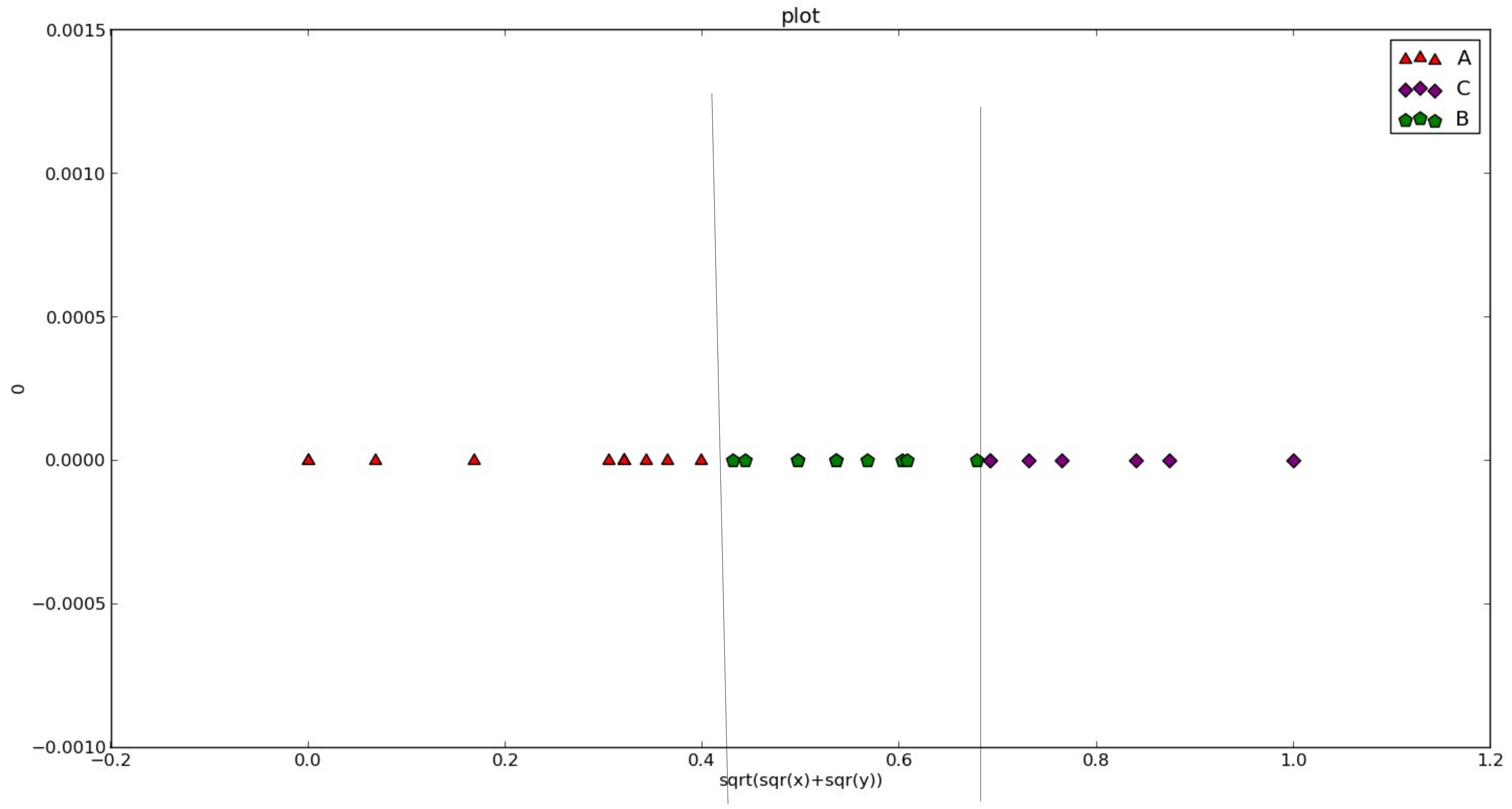
Go Frendi Gunawan (5111201033)

Ruang Fitur

x	y	kelas
0.3	0.5	A
0.4	0.9	B
0.6	0.2	A
0.9	1.0	C
...	...	...
1.0	0.3	C
0.8	0.2	B



Ekstraksi Fitur

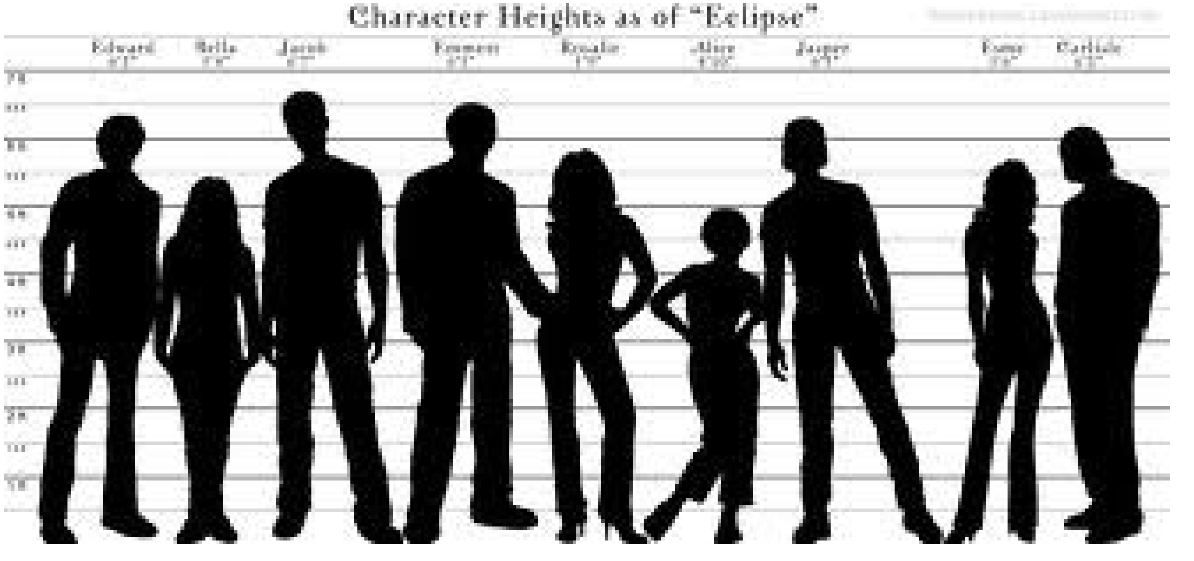


Penelitian Sebelumnya

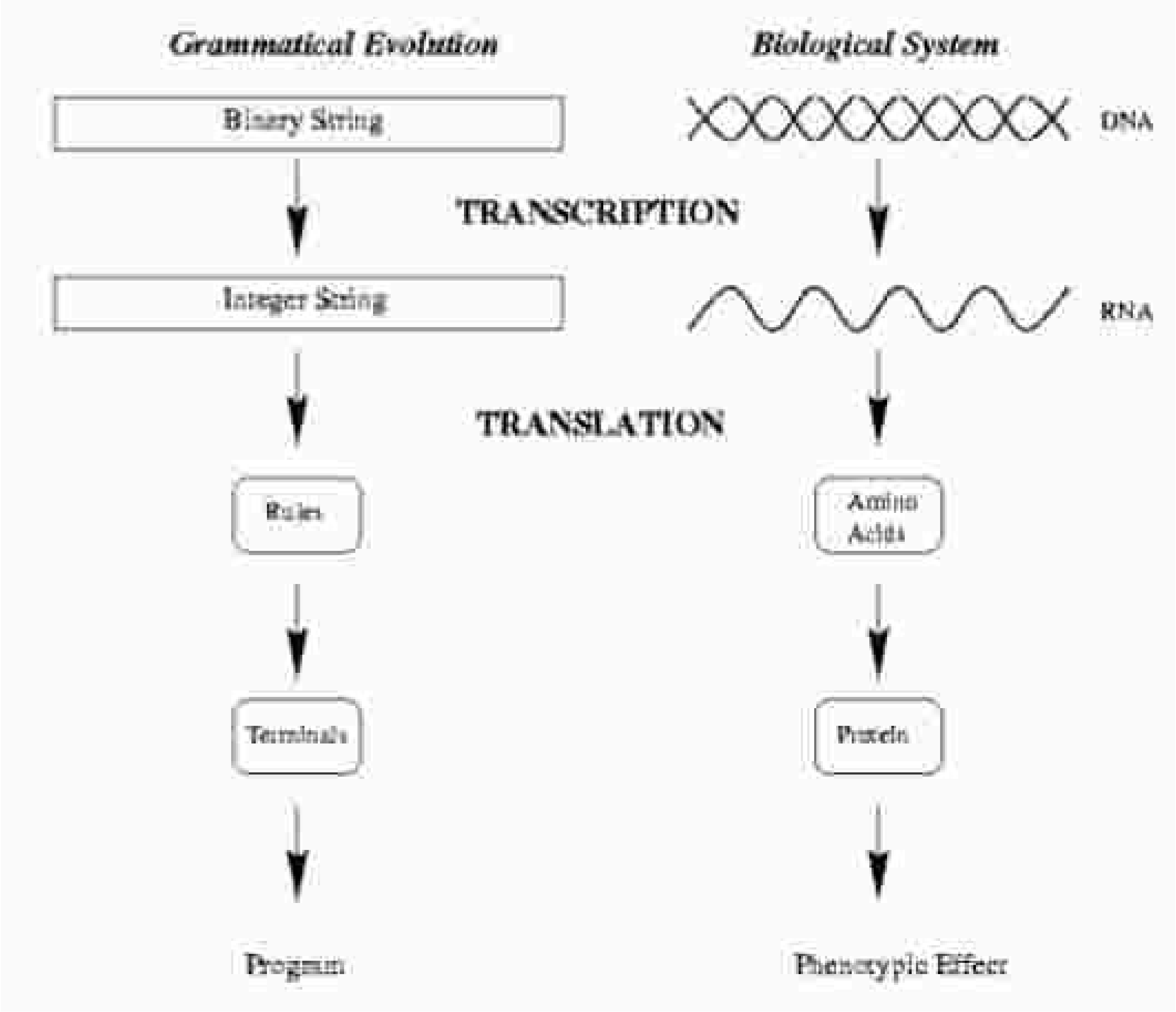
- [1] Gunawan G. F., Gosaria S, Arifin A. Z. (2012). “Grammatical Evolution For Feature Extraction In Local Thresholding Problem”, Jurnal Ilmu Komputer dan Informasi, Vol 5, No 2 (2012)
- [2] Harper R., Blair A. (2006). “Dynamically Define Functions in Grammatical Evolution”, IEEE Congress of Evolutionary Computation, July 16-21, 2006
- [3] Gavrilis D., Tsoulous I. G., Georgoulas G., Glavas E. (2005). "Classification of Fetal Heart Rate Using Grammatical Evolution", IEEE Workshop on Signal Processing Systems Design and Implementation, 2005.
- [4] Gavrilis D., Tsoulous I. G., Dermatas E. (2008). "Selecting and Constructing Features Using Grammatical Evolution", Journal Pattern Recognition Letters Volume 29 Issue 9, July, 2008 Pages 1358-1365 .
- [5] Guo L., Rivero D., Dorado J., Munteanu C. R., Pazos A. (2011). "Automatic feature extraction using genetic programming: An application to epileptic EEG classification ", Expert Systems with Applications 38 Pages 10425-10436
- [6] Li B., Zhang P.Y., Tian H., Mi S.S., Liu D.S., Ruo G.Q. (2011). "A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox", Expert Systems with Applications 38 Pages 10000-10009
- [7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. , Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P. , Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). "Scikit-learn: Machine Learning in Python", Journal of Machine Learning Research Vol. 12 Pages 2825-2830

Penelitian yang Dilakukan

- Classifier: Decision Tree
- 5 Feature Extractor untuk dibandingkan:
  - GA Select
  - GE Gavrilis
  - GE Global
  - **GE Multi**
  - **GE Tatami**



Grammatical Evolution (Overview)



Sumber : Ryan Conor, Gecco 2006 Grammatical Evolution Tutorial

Grammatical Evolution (Detail)

- Given the individual

220	203	51	123	2	45

....what will happen?

- $<$expr$>$ has 4 production rules to choose from

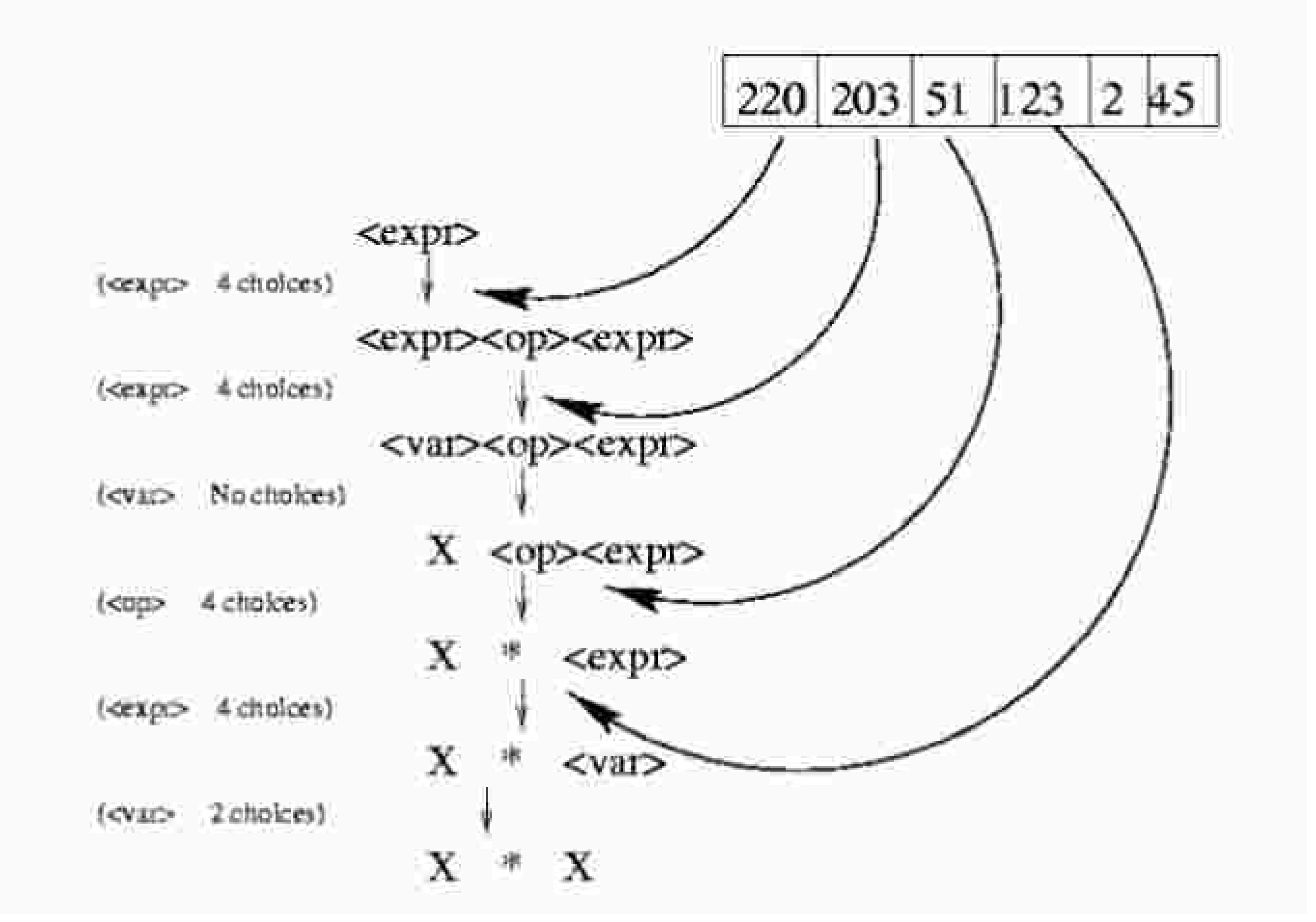
```
(1) <expr> ::= <expr> <op> <expr> (A)
             | ( <expr> <op> <expr> ) (B)
             | <pre-op> ( <expr> ) (C)
             | <var> (D)
```

  - Taking first codon 220 we get 220 $MOD\ 4 = 0$
  - Gives $<\underline{\text{expr}}>< op >< expr >$

- Next choice for the first $<\underline{\text{expr}}>$
  - Taking next codon 203 we get 203 $MOD\ 4 = 3$
  - Gives $<\underline{\text{var}}>< op >< expr >$

Sumber : Ryan Conor, Gecco 2006 Grammatical Evolution Tutorial

Grammatical Evolution (Detail)



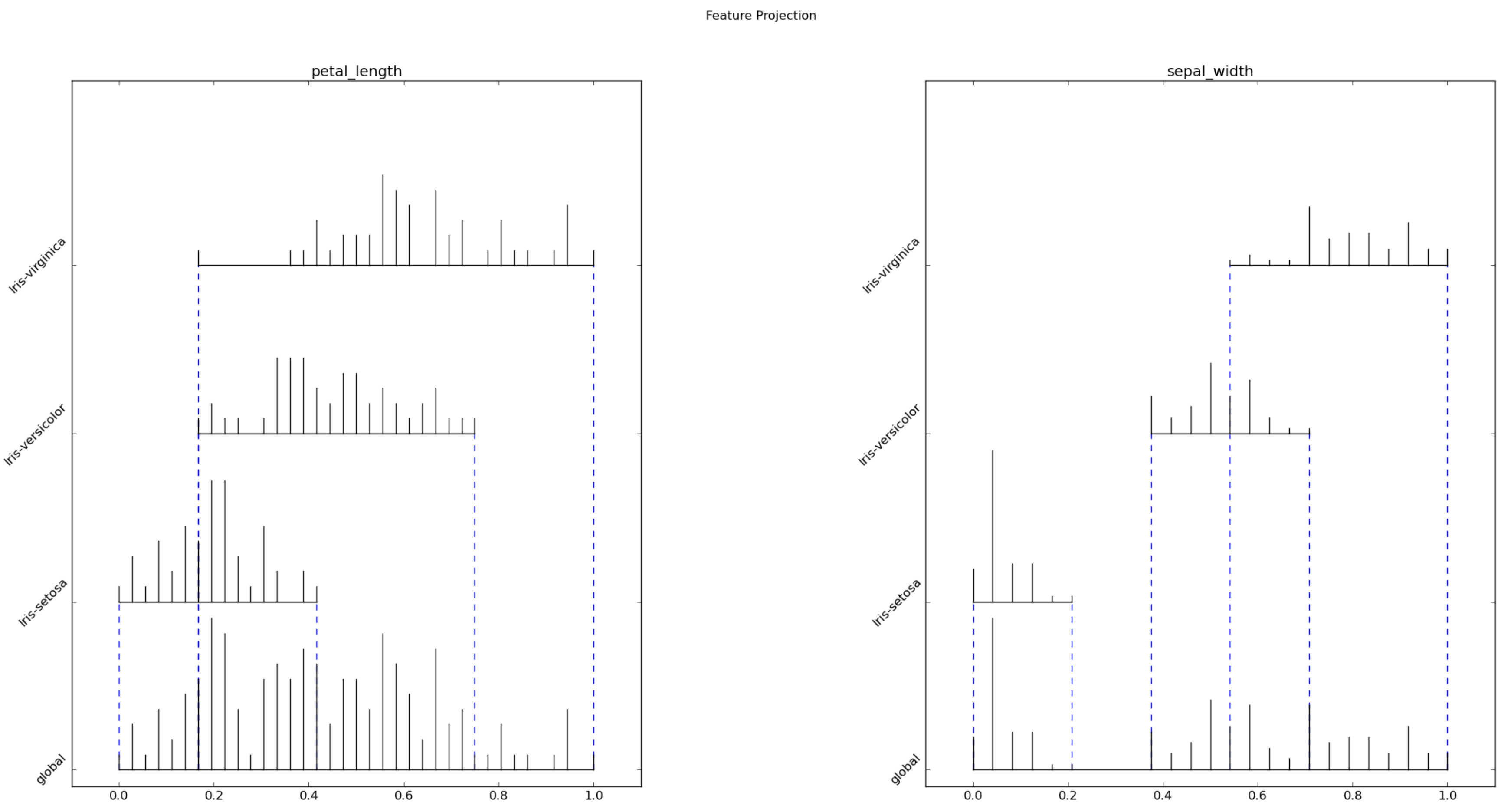
Sumber : Ryan Conor, Gecco 2006 Grammatical Evolution Tutorial

GA Select

- Algoritma Genetik
- Individu di-evolusikan menjadi **subset fitur original**
- Fitness Value = Akurasi Classifier menggunakan subset fitur original
- Menilai keakuratan subset fitur tanpa modifikasi.
  - Misal: Fitur original = {x,y,z}
  - Subset fitur yang dihasilkan:
    - {x}
    - {x,y}
    - {x,z}
    - {y}
    - {y,z}
    - {z}

Fitur yang Dihasilkan GA Select

- **Fitur original(4)** : sepal_length, sepal_width, petal_length, petal_width
- **Fitur yang dihasilkan (2)** : petal_length, sepal_width

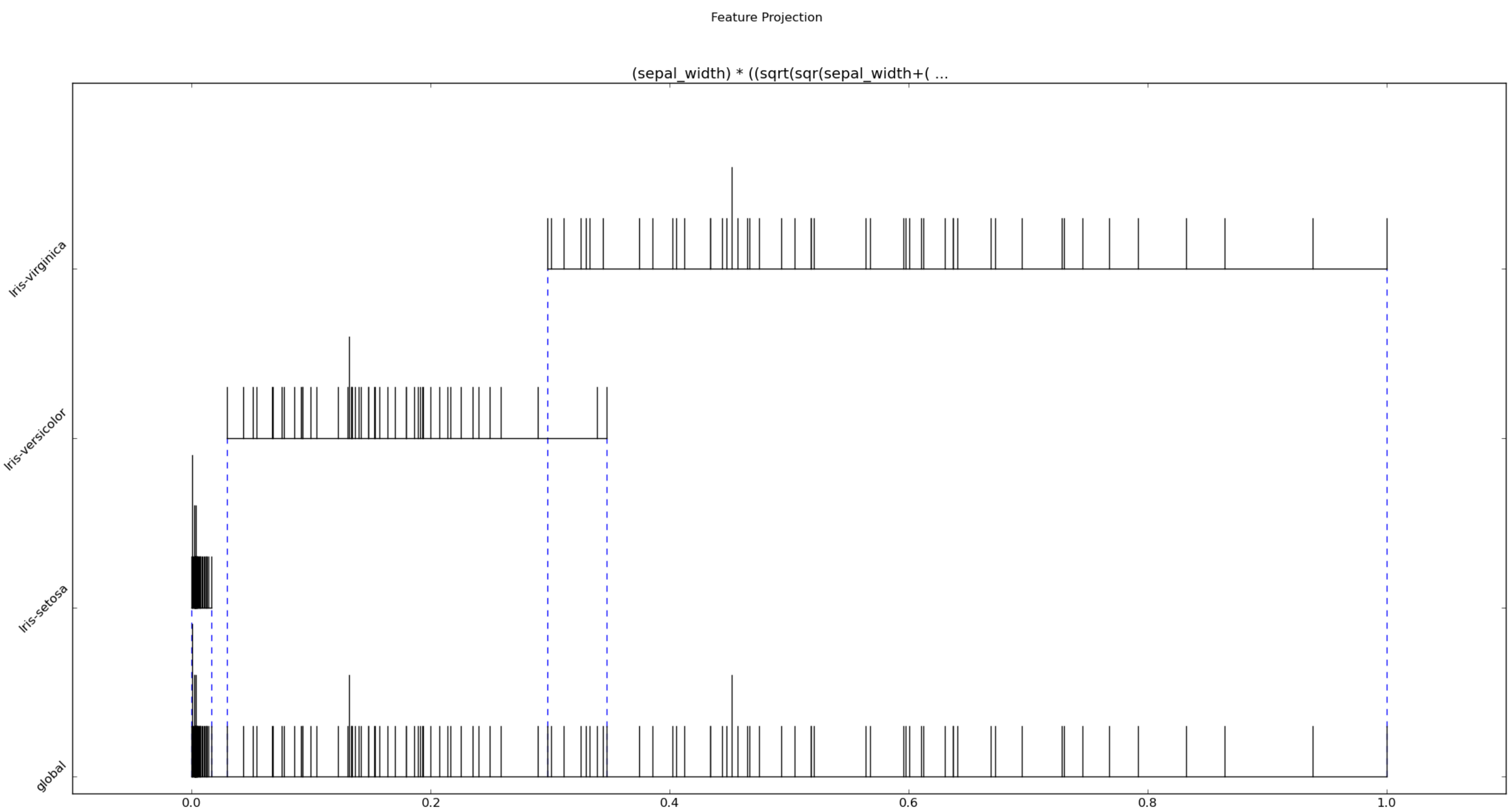


GE Gavrilis

- Grammatical Evolution
- Individu di-evolusikan menjadi **1 set fitur baru**
- Dihasilkan sejumlah fitur secara acak, namun tidak semuanya relevan
- Fitness Value = Akurasi Classifier menggunakan set fitur baru
- Gavrilis D., et al. (2005)

Contoh Fitur yang dihasilkan GE Gavrilis

- **Fitur original (4):** sepal_length, sepal_width, petal_length, petal_width
- **Fitur yang dihasilkan (1):** (sepal_width) * ((sqrt(sqr(sepal_width+(sepal_length) * ((sepal_width) / (sepal_width)))/2)) * (abs(((sepal_length) - (petal_length)) + (sepal_length))))

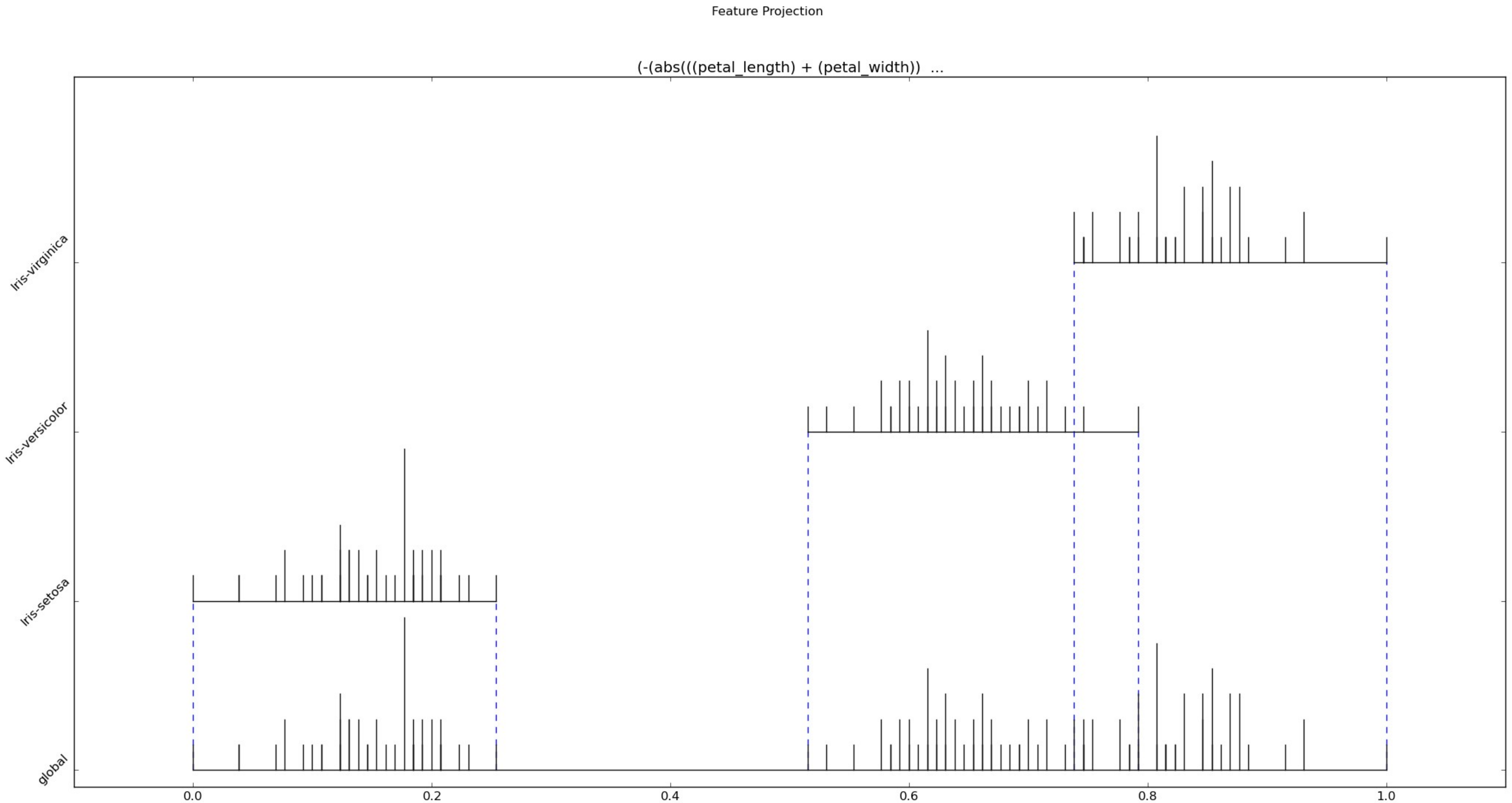


GE Global

- Grammatical Evolution
- Menghasilkan **1 fitur baru**
- Fitness Value = Akurasi Classifier menggunakan 1 fitur baru

Contoh Fitur yang dihasilkan GE Global

- **Fitur original :** sepal_length, sepal_width, petal_length, petal_width
- **Fitur yang dihasilkan (1):** (-(abs(((petal_length) + (petal_width)) - (sepal_width)))) + ((sepal_length) + (sepal_length))

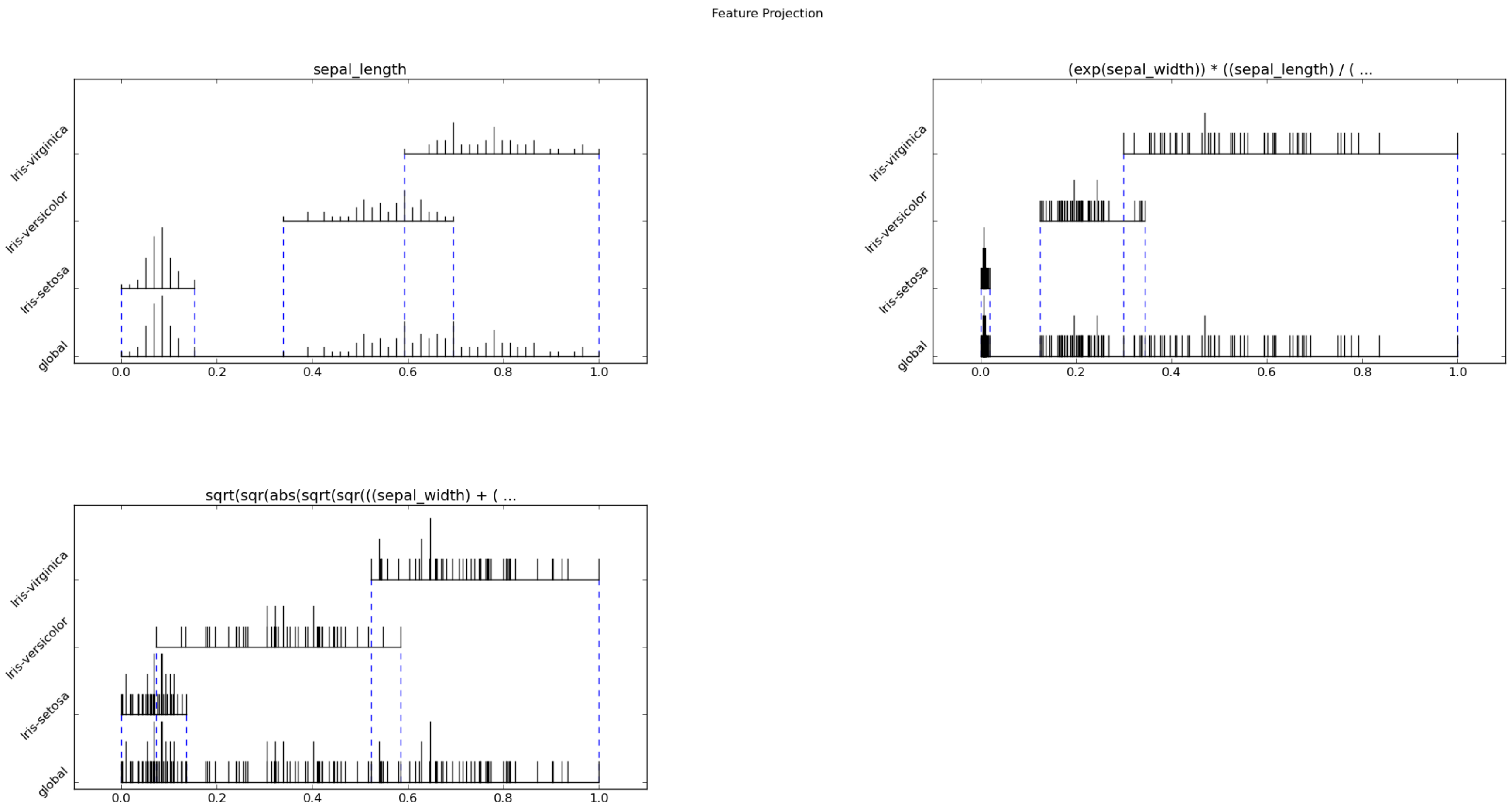


GE Multi

- Grammatical Evolution
- Menghasilkan **n fitur baru**, n = jumlah kelas
- Fitness Value = Akurasi Classifier menggunakan n fitur baru

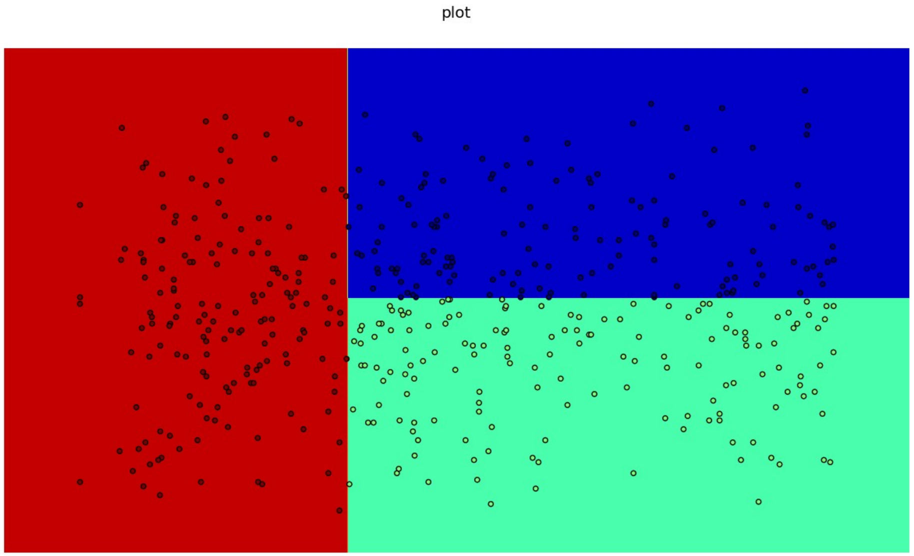
Contoh Fitur yang dihasilkan GE Multi

- **Fitur original :** sepal_length, sepal_width, petal_length, petal_width
- **Fitur yang dihasilkan (3):** sepal_length, (exp(sepal_width)) * ((sepal_length) / (petal_width)), sqrt(sqr(abs(sqrt(sqr(((sepal_width) + ((sepal_width) - (sqrt(sqr((petal_length) - (sepal_length)+sepal_width)/2)))) - (petal_width)+petal_width)/2))+sepal_length)/2)

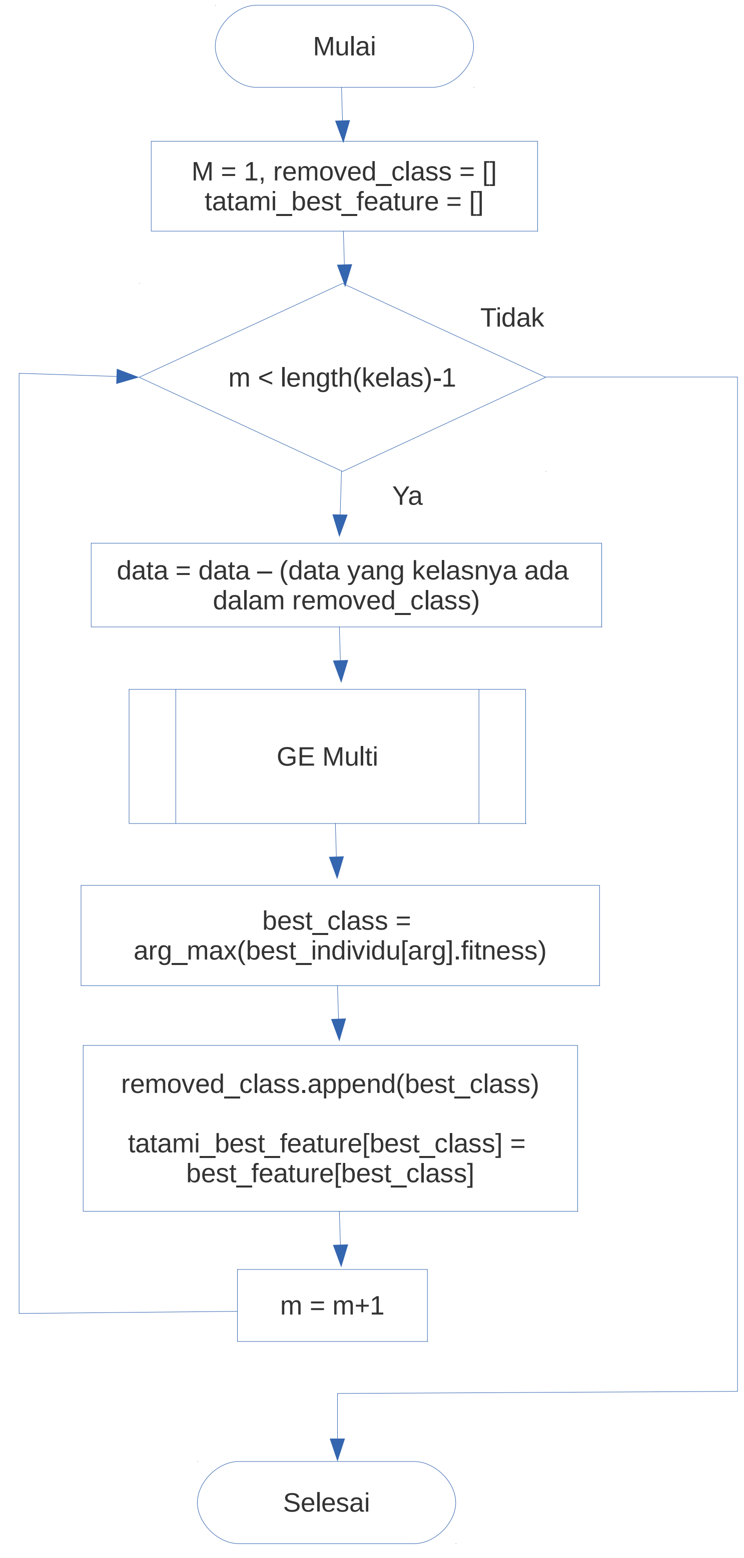


GE Tatami

- Grammatical Evolution
- Menghasilkan **n-1 fitur baru**, n = jumlah kelas
- Fitness Value = Akurasi Classifier menggunakan n-1 fitur baru

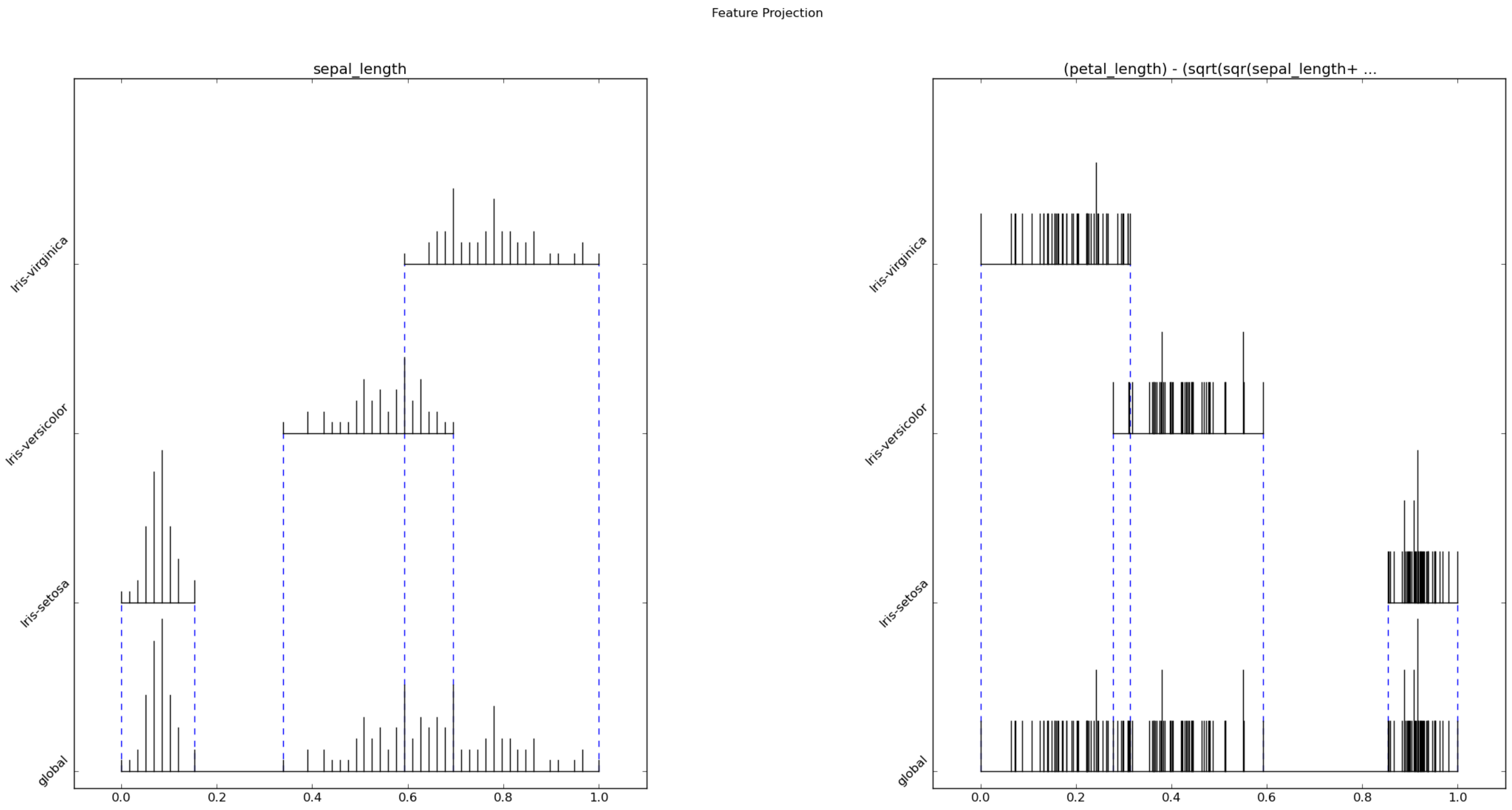


Flowchart GE Tatami



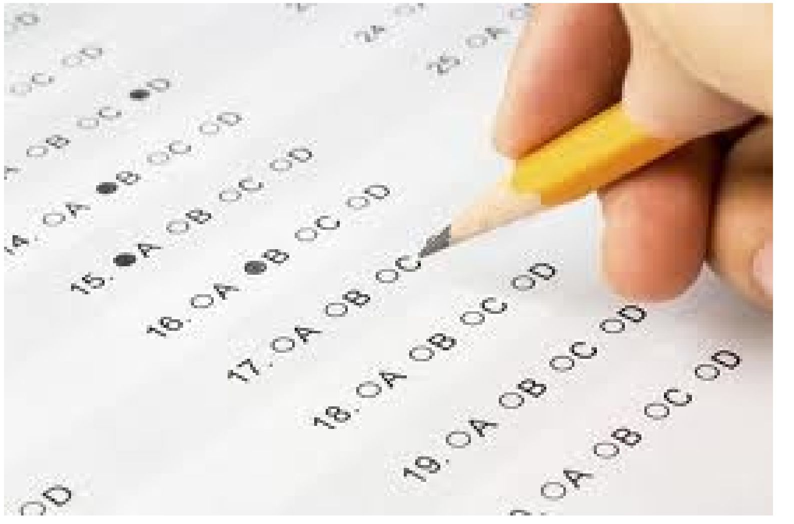
Contoh Fitur yang Dihasilkan GE Tatami

- **Fitur original :** sepal_length, sepal_width, petal_length, petal_width
- **Fitur yang dihasilkan (2):** sepal_length, (petal_length) - (sqrt(sqr(sepal_length+sqrt(sqr(abs(petal_length)+(sqrt(sqr(petal_length+petal_width)/2)) - (abs(-((sepal_width) - (petal_width)))))/2))/2))



Dataset Pengujian

- Dataset
  - Sintesis 01
  - Sintesis 02
  - Sintesis 03
  - Iris (UCI Machine Learning)
  - E Coli (UCI Machine Learning)
  - Balanced Scale (UCI Machine Learning)
- 5 Fold Cross Validation



Hasil Pengujian Dataset Sintesis 01

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
Whole	**Train**	73.04	3	77.83	1	99.35	3	***100.0***	2	85.65	3
	Test	73.04		77.83		99.35		***100.0***		85.65	
	Total	73.04		77.83		99.35		***100.0***		85.65	
Fold 1	**Train**	75.95	3	78.92	1	**100.0**	3	***100.0***	2	84.59	61
	Test	67.78		35.56		76.67		*81.11*		**83.33**	
	Total	74.35		70.43		95.43		***96.3***		84.35	
Fold 2	**Train**	73.78	3	78.92	1	**100.0**	3	***100.0***	2	**100.0**	48
	Test	70.0		36.67		74.44		*65.56*		**80.0**	
	Total	73.04		70.65		95.0		*93.26*		**96.09**	
Fold 3	**Train**	71.62	3	77.57	1	**100.0**	3	***100.0***	2	85.14	3
	Test	75.56		25.56		**86.67**		***86.67***		**86.67**	
	Total	72.39		67.39		**97.39**		***97.39***		85.43	
Fold 4	**Train**	73.51	3	78.65	1	**100.0**	3	***100.0***	2	85.41	3
	Test	74.44		36.67		**77.78**		***77.78***		76.67	
	Total	73.7		70.43		**95.65**		***95.65***		83.7	
Fold 5	**Train**	75.68	3	81.08	1	**100.0**	3	***100.0***	2	87.03	2
	Testing	70.0		42.22		**94.44**		*82.22*		72.22	
	Total	74.57		73.48		**98.91**		*96.52*		84.13	

Hasil Pengujian Dataset Sintesis 02

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
Whole	Train	77.98	4	70.8	1	**100.0**	4	***100.0***	3	90.38	12
	Test	77.98		70.8		**100.0**		***100.0***		90.38	
	Total	77.98		70.8		**100.0**		***100.0***		90.38	
Fold 1	Train	78.66	4	73.17	1	99.39	4	***100.0***	3	89.84	12
	Test	**76.03**		33.06		46.28		*62.81*		72.73	
	Total	78.14		65.25		88.91		***92.66***		86.46	
Fold 2	Train	76.42	4	70.53	1	**100.0**	4	***100.0***	3	89.43	12
	Test	77.69		27.27		72.73		*74.38*		**85.95**	
	Total	76.67		61.99		94.62		***94.94***		88.74	
Fold 3	Train	79.67	4	71.75	1	99.39	3	***100.0***	3	90.04	12
	Test	68.6		34.71		64.46		***83.47***		68.6	
	Total	77.49		64.44		92.5		***96.74***		85.81	
Fold 4	Train	79.07	4	70.73	1	**100.0**	4	***100.0***	3	90.24	12
	Test	72.73		40.5		71.07		*61.16*		**80.17**	
	Total	77.81		64.76		**94.29**		*92.33*		88.25	
Fold 5	Train	78.05	4	71.75	1	100.0	4	***100.0***	3	86.99	12
	Test	73.55		32.23		83.47		***94.21***		70.25	
	Total	77.16		63.95		96.74		***98.86***		83.69	

Hasil Pengujian Dataset Sintesis 03

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
Whole	**Train**	72.5	6	67.0	1	98.25	5	***100.0***	4	80.75	6
	Test	72.5		67.0		98.25		***100.0***		80.75	
	Total	72.5		67.0		98.25		***100.0***		80.75	
Fold 1	**Train**	74.45	6	67.91	1	99.07	5	***100.0***	4	82.55	6
	Test	26.58		27.85		45.57		***63.29***		32.91	
	Total	65.0		60.0		88.5		***92.75***		72.75	
Fold 2	**Train**	72.9	6	66.98	1	100.0	5	***100.0***	4	86.29	47
	Test	62.03		27.85		68.35		***94.94***		69.62	
	Total	70.75		59.25		93.75		***99.0***		83.0	
Fold 3	**Train**	71.34	6	69.47	1	99.69	5	***100.0***	4	83.18	6
	Test	29.11		24.05		55.7		***65.82***		46.84	
	Total	63.0		60.5		91.0		***93.25***		76.0	
Fold 4	**Train**	72.9	4	66.98	1	98.44	5	***100.0***	4	73.52	2
	Test	27.85		26.58		50.63		***78.48***		25.32	
	Total	64.0		59.0		89.0		***95.75***		64.0	
Fold 5	**Train**	72.27	6	68.85	1	99.38	5	***100.0***	4	85.67	47
	Test	25.32		48.1		59.49		***63.29***		49.37	
	Total	63.0		64.75		91.5		***92.75***		78.5	

Hasil Pengujian Dataset Iris

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
Whole	Train	96.0	2	**98.67**	1	**98.67**	3	***98.67***	2	**98.67**	1
	Test	96.0		**98.67**		**98.67**		***98.67***		**98.67**	
	Total	96.0		**98.67**		**98.67**		***98.67***		**98.67**	
Fold 1	Train	96.67	2	**99.17**	1	**99.17**	3	*98.33*	2	**99.17**	1
	Test	86.67		**96.67**		**96.67**		***96.67***		**96.67**	
	Total	94.67		**98.67**		**98.67**		*98.0*		**98.67**	
Fold 2	Train	95.83	2	**100.0**	1	99.17	3	*99.17*	2	**100.0**	3
	Test	96.67		**96.67**		83.33		*66.67*		**96.67**	
	Total	96.0		**99.33**		96.0		*92.67*		**99.33**	
Fold 3	Train	96.67	2	98.33	1	98.33	3	***99.17***	2	98.33	1
	Test	93.33		76.67		**100.0**		***100.0***		**100.0**	
	Total	96.0		94.0		98.67		***99.33***		98.67	
Fold 4	Train	95.83	2	99.17	1	**99.17**	3	***99.17***	2	**99.17**	1
	Test	96.67		93.33		**96.67**		***96.67***		**96.67**	
	Total	96.0		98.0		**98.67**		***98.67***		**98.67**	
Fold 5	Train	96.67	2	98.33	1	98.33	3	***99.17***	2	**99.17**	3
	Testing	93.33		93.33		93.33		***96.67***		**96.67**	
	Total	96.0		97.33		97.33		***98.67***		**98.67**	

Hasil Pengujian Dataset E Coli

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
Whole	**Train**	97.02	7	84.52	1	96.73	8	***97.62***	7	97.02	12
	Test	97.02		84.52		96.73		***97.62***		97.02	
	Total	97.02		84.52		96.73		***97.62***		97.02	
Fold 1	**Train**	97.42	6	87.45	1	**98.89**	8	*96.31*	7	97.79	12
	Test	**73.85**		58.46		53.85		*53.85*		69.23	
	Total	**92.86**		81.85		90.18		*88.1*		92.26	
Fold 2	**Train**	96.68	5	86.72	1	97.79	8	***98.52***	7	97.79	12
	Test	**78.46**		63.08		49.23		*1.54*		58.46	
	Total	**93.15**		82,14		88.39		*79.76*		90.18	
Fold 3	**Train**	97.79	7	89.3	1	**99.26**	8	*97.05*	7	98.52	26
	Test	70.77		53.85		**80.0**		*69.23*		53.85	
	Total	92.56		82.44		**95.54**		*91.67*		89.88	
Fold 4	**Train**	96.31	5	87.08	1	**98.52**	8	*97.42*	7	98.15	12
	Test	**73.85**		69.23		63.08		*43.08*		56.92	
	Total	**91.96**		83.63		91.67		*86.9*		90.18	
Fold 5	**Train**	97.05	4	86.72	1	98.15	8	*96.31*	7	**98.89**	18
	Test	**75.38**		61.54		38.46		*38.46*		67.69	
	Total	**92.86**		81.85		86.61		*85.12*		**92.86**	

Hasil Pengujian Dataset Balanced Scale

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
Whole	Train	70.88	3	84.8	1	***91.68***	1	***91.68***	2	82.56	9
	Test	70.88		84.8		***91.68***		***91.68***		82.56	
	Total	70.88		84.8		***91.68***		***91.68***		82.56	
Fold 1	Train	70.92	3	**100.0**	1	**100.0**	1	*92.03*	2	81.08	4
	Test	70.73		**91.87**		**91.87**		*85.37*		81.3	
	Total	70.88		**98.4**		**98.4**		*90.72*		81.12	
Fold 2	Train	71.91	3	85.46	1	92.23	1	***92.63***	2	83.86	9
	Test	66.67		69.92		**89.43**		*82.93*		78.86	
	Total	70.88		82.4		**91.68**		*90.72*		82.88	
Fold 3	Train	70.52	3	90.04	1	**99.0**	2	*92.03*	2	83.67	126
	Test	72.36		66.67		**85.37**		*71.54*		81.3	
	Total	70.88		85.44		**96.32**		*88.0*		83.2	
Fold 4	Train	72.51	3	86.65	1	**100.0**	1	*91.83*	2	82.27	2
	Test	68.29		78.05		73.98		***91.06***		77.24	
	Total	71.68		84.96		**94.88**		*91.68*		81.28	
Fold 5	Train	71.12	3	84.66	1	94.62	3	***100.0***	2	82.87	1
	Test	69.92		**82.93**		57.72		*66.67*		51.22	
	Total	70.88		84.32		87.36		***93.44***		76.64	

Rata Rata Hasil Pengujian

Experiment		GA Select Feature		GE Global		GE Multi		GE Tatami Multi		GE Gavrilis	
		Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features	Accuracy (%)	Features
iris.data	Train	96.28	2	98.95	1	98.81	3	98.95	2	**99.08**	1
	Test	93.78		92.56		94.78		92.56		**97.56**	
	Total	95.78		97.67		98.0		97.67		**98.78**	
balance-scale.data	Train	71.31	3	88.6	1	**96.25**	1	93.37	2	82.72	25
	Test	69.81		79.04		**81.68**		81.54		75.41	
	Total	71.01		86.72		**93.39**		91.04		81.28	
ecoli.edit	Train	97.05	5	86.96	1	**98.22**	8	97.2	7	98.03	15
	Test	**78.22**		65.11		63.56		50.63		67.2	
	Total	**93.4**		82.74		91.52		88.2		92.06	
synthesis_01	Train	73.93	3	78.83	1	99.89	3	**100.0**	2	87.97	20
	Test	71.8		42.42		**84.89**		82.22		80.76	
	Total	73.52		71.7		**96.95**		96.52		86.56	
synthesis_02	Train	78.31	4	71.45	1	99.8	3	**100.0**	3	89.49	12
	Test	74.43		39.76		73.0		**79.34**		78.01	
	Total	77.54		65.2		94.51		**95.92**		87.22	
synthesis_03	Train	72.73	5	67.87	1	99.14	5	**100.0**	4	81.99	19
	Test	40.57		36.9		63.0		**77.64**		50.8	
	Total	66.38		61.75		92.0		**95.58**		75.83	
All	Train	81.6	3	82.11	1	**98.69**	3	98.25	3	89.88	15
	Test	71.49		59.3		76.82		**77.32**		74.96	
	Total	79.6		77.63		**94.4**		94.15		86.96	

Kesimpulan

- Secara umum GE Tatami & GE Multi dapat menciptakan set fitur yang dapat meningkatkan akurasi decision tree.
- Penggunaan SVM sebagai classifier memberikan hasil yang buruk.
- GE Tatami memberikan persyaratan yang lebih mudah dipenuhi dibandingkan GE Multi & GE Global.
- Jika pada langkah pertama GE Tatami gagal mencari fitur terbaik, maka langkah selanjutnya tidak akan bisa memperbaiki akurasi.

Saran

- Penggabungan GE Multi & GE Tatami diharapkan dapat memberikan hasil yang lebih baik.
- Optimalisasi fitness function dapat meningkatkan kecepatan komputasi

Thank you ...

ありがとう ...

谢谢 ...

Terima kasih ...