

# ອຸນວຍກາສແໜ່ງດາວ ມີນບດແລະກັບປະໂຫຍດ

ເມືອງເຄມບະ





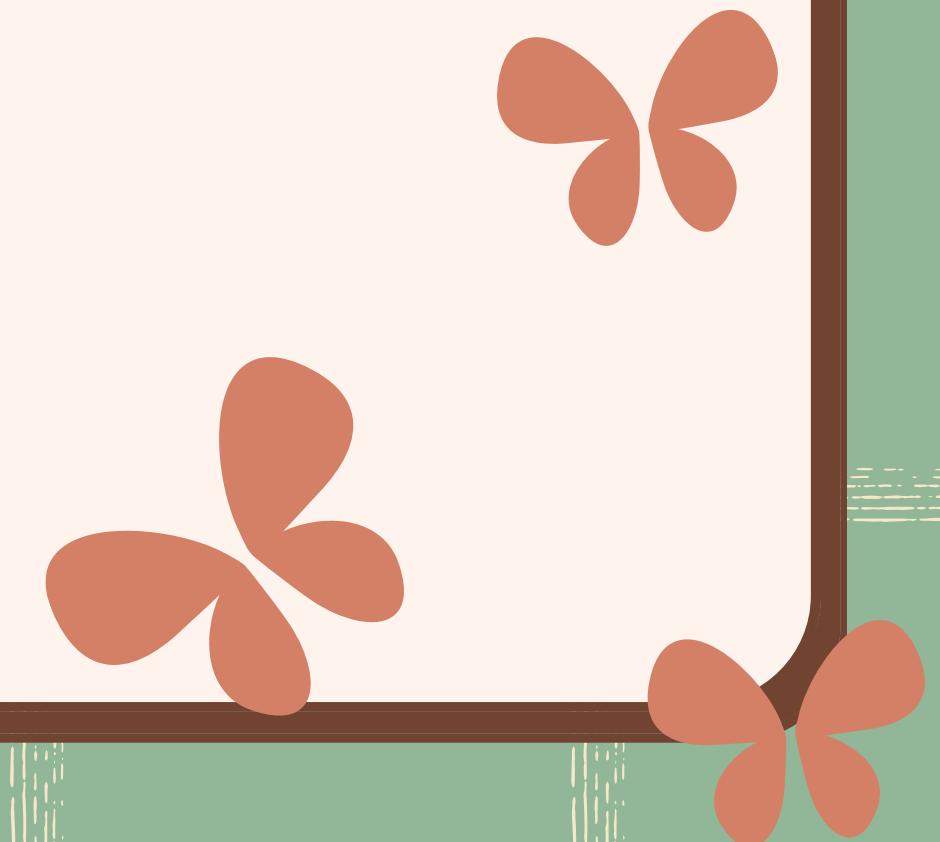
# การสำรวจข้อมูล เลือก DATA



STD_ENCODE_ID	IC3_MODULE_NAME	IC3_RESULT	IC3_EXAM_TIMEUSED	IC3_EXAM_YEAR	STD_FACULTYNAME_THAI	CLASS_ENROLLSEAT	CLASS_ACADEMIC_YEAR	CLASS_SEMESTER	ONLINE_ASSIGNMENT_SUBMISSION_FREQUENCY
student12215	IC3 GS5 - Computing Fundamentals	Pass	1455	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12215	IC3 GS5 - Living Online	Pass	856	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12215	IC3 GS5 - Key Applications	Fail	3000	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12216	IC3 GS5 - Computing Fundamentals	Pass	1751	2021	วิศวกรรมศาสตร์	7	2020	2	High
student12216	IC3 GS5 - Living Online	Pass	1616	2021	วิศวกรรมศาสตร์	7	2020	2	High
student12216	IC3 GS5 - Key Applications	Pass	2925	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12217	IC3 GS5 - Key Applications	Pass	2221	2021	วิศวกรรมศาสตร์	7	2020	2	High
student12217	IC3 GS5 - Computing Fundamentals	Pass	1094	2021	วิศวกรรมศาสตร์	7	2020	2	High
student12217	IC3 GS5 - Living Online	Pass	1512	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12217	IC3 GS5 - Key Applications	Fail	3000	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12219	IC3 GS5 - Computing Fundamentals	Pass	1578	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12219	IC3 GS5 - Living Online	Pass	1361	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12219	IC3 GS5 - Key Applications	Pass	3000	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12220	IC3 GS5 - Computing Fundamentals	Pass	1535	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12220	IC3 GS5 - Living Online	Pass	1917	2021	วิศวกรรมศาสตร์	7	2020	2	Medium
student12220	IC3 GS5 - Kev Annlications	Pass	2979	2021	วิศวกรรมศาสตร์	7	2020	2	High

คอลัมบ์ที่ใช้ในการสร้างโมเดล

STD\_ENCODE\_ID  
IC3\_MODULE\_NAME  
IC3\_RESULT  
IC3\_EXAM\_TIMEUSED  
ASS\_ACADEMIC\_YEAR  
CLASS\_SEMESTER  
ONLINE\_ASSIGNMENT\_SUBMISSION\_FREQUENCY

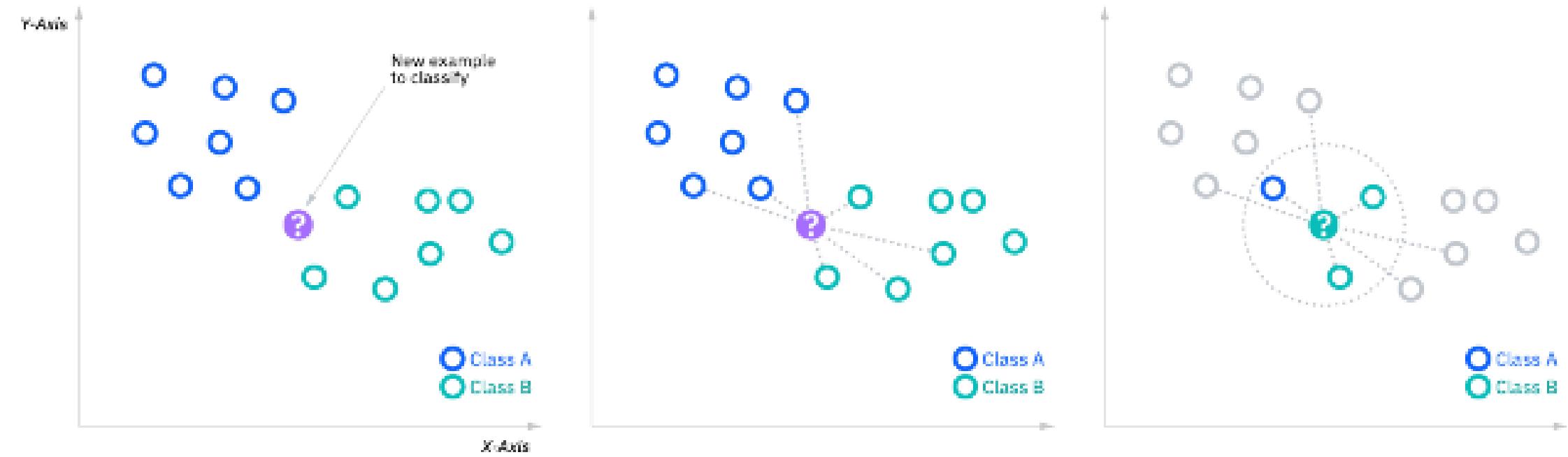




# ກຸບລົງຂອງ KNN



K-Nearest Neighbors ນີ້ແມ່ນເຄີຍກວ່າ KNN ນີ້ວ່າ k-NN ເປັນວິທີການແບ່ງດຸລສສໍາຫຼັບໃຈໜັດມົວນຸ່ມື້ຂ້ອມູລ (Classification) ໂດຍມີໜັກການນຳຂ້ອມູລວິ່ນໆມາເປົ້າຍບເກີຍບກັບຕົວຂ້ອມູລທີ່ສັນໃຈ ວ່າມີຄວາມໃກລ້າເຕີຍງກັນມາກແຄ່ໃໝ່ ນາກຂ້ອມູລທີ່ສັນໃຈອູ່ໃກລ້າກັບຂ້ອມູລໃດມາກທີ່ສຸດ ຮະບບຈະໃໝ່ຄຳຕອບເປັນເໜືອນຄຳຕອບຂອງຂ້ອມູລທີ່ອູ່ໃກລ້າທີ່ສຸດ

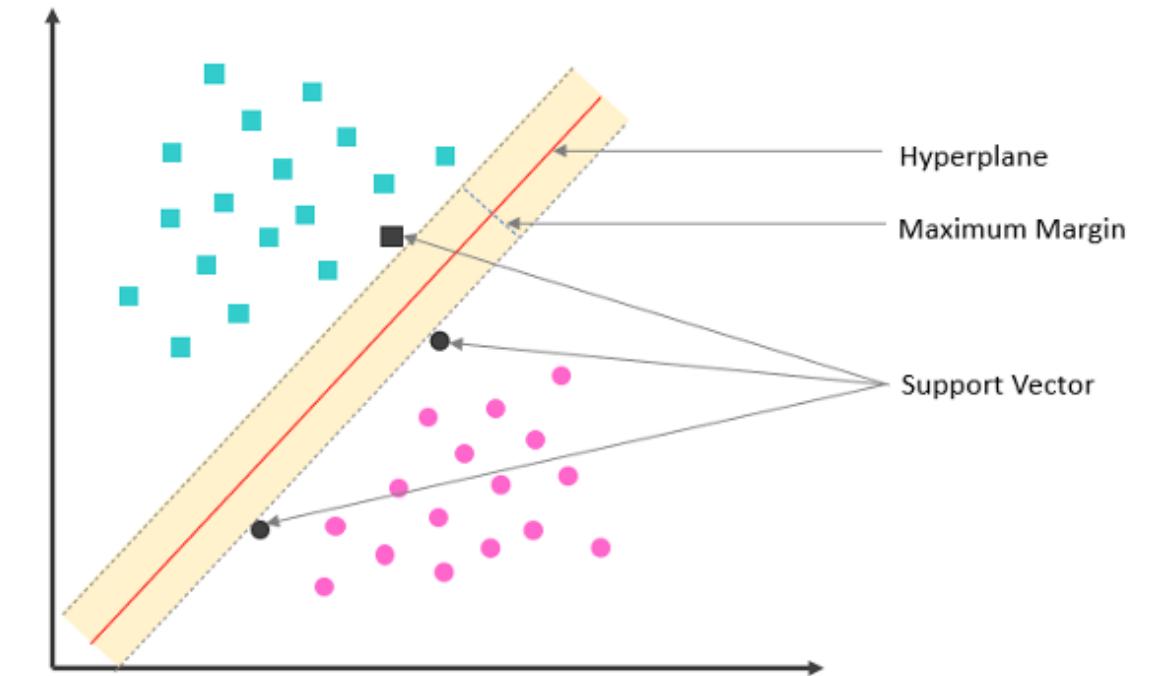
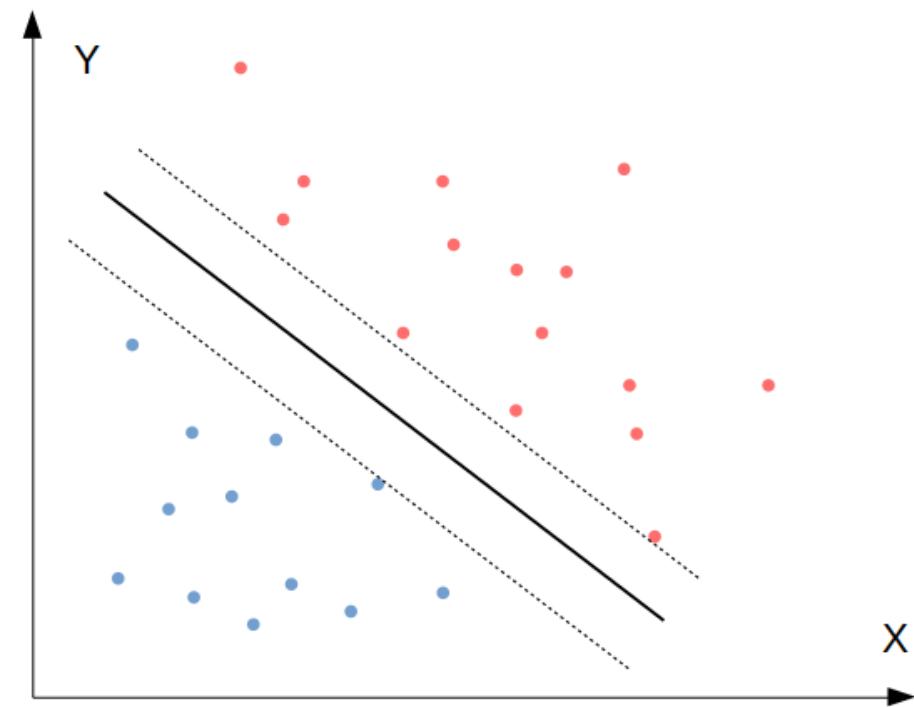




# ການຮັບສ້າງ SVM

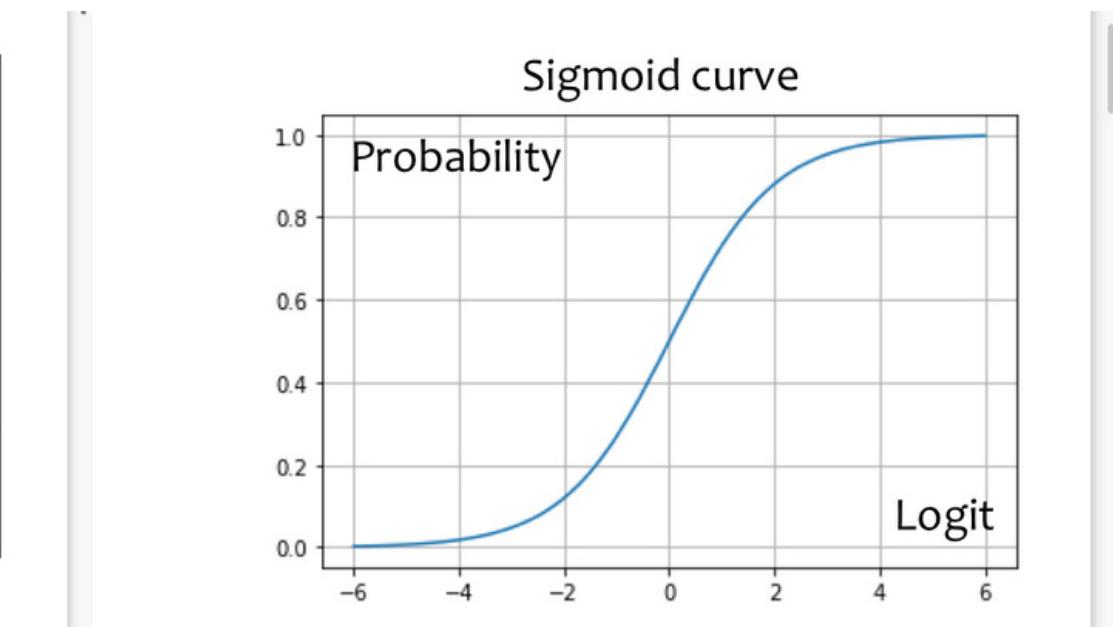
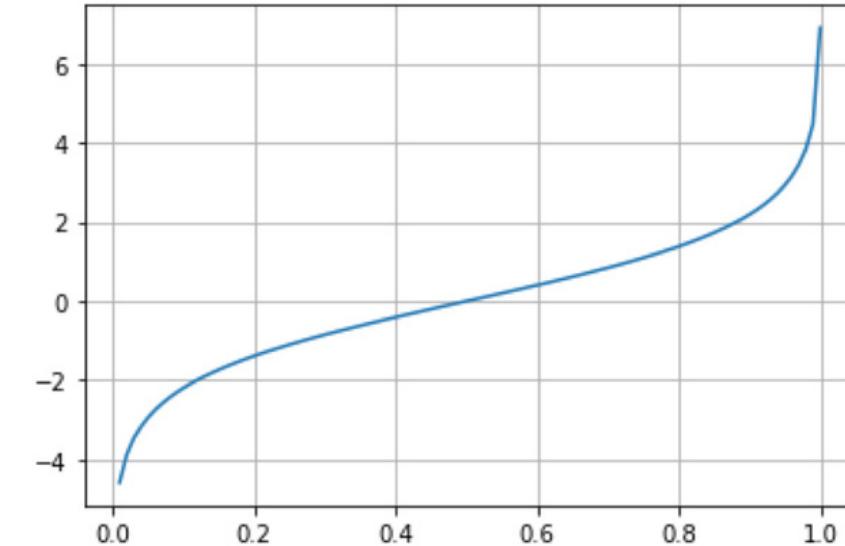


SVM (Support Vector Machine) ເປັນຕົວຈຳແນກເຊີງເສັ້ນ (Linear Classifier) ແບບ ໄບນາຣີ (Binary) (ແບ່ງແຍກຂໍ້ອມູລໄດ້ 2 ປະເກດ) ຂໍອໄດ້ເປີຍບ່ອງ SVM ດື່ມມີປະສິກິພາພ ໃນການຈຳແນກຂໍ້ອມູລທີ່ມີມີຕິຈຳນວນຫຼາຍໄດ້ ນອກຈາກນີ້ການໃໝ່ຝຶງກໍ່ຈັນເຄວົ່ນ (Kernel Function) ຍັງຊ່ວຍໃໝ່ສາມາຮັດຈຳແນກຂໍ້ອມູລທີ່ມີຄວາມຄລຸມເຄຣີວໄດ້ຢ່າງມີປະສິກິພາພ



# ☆ 🔒 ກຖະງົດຂອງ Logistic Regression Q ⌂≡

ກຖະງົດ Logistic Regression ເປັນໜຶ່ງໃນວິລກອຣີກິມໃນກາຮເຮືຍນຮູ້ຂອງເຄື່ອງ (Machine Learning) ກີ່ໃຊ້ໃນກາຮຈໍາແນກປະເກົາຂໍ້ມູນ ໂດຍຕົວແປຣຕົ້ນ (independent variable) ກີ່ໃຊ້ໃນກາຮກໍານາຍຫຼືວິຈໍາແນກຈະເປັນຕົວແປຣກີ່ມີລັກບະນະທີ່ເປັນບຮຈຸໄດ້ເປັນຕົວແປຣເດືອວ (binary) ພົບສາມາຄແບ່ງເປັນກລຸ່ມໄດ້ເປັນຕັ້ງແຕ່ 2 ກລຸ່ມ (multinomial) ພົບມາກກວ່ານັ້ນແຕ່ໃນກຣນີກີ່ມີກລຸ່ມເປັນ 2 ກລຸ່ມ ເຮົາເຮີຍກວ່າ Binary Logistic Regression





# ກຸບລົງຂອງ Naive Bayes



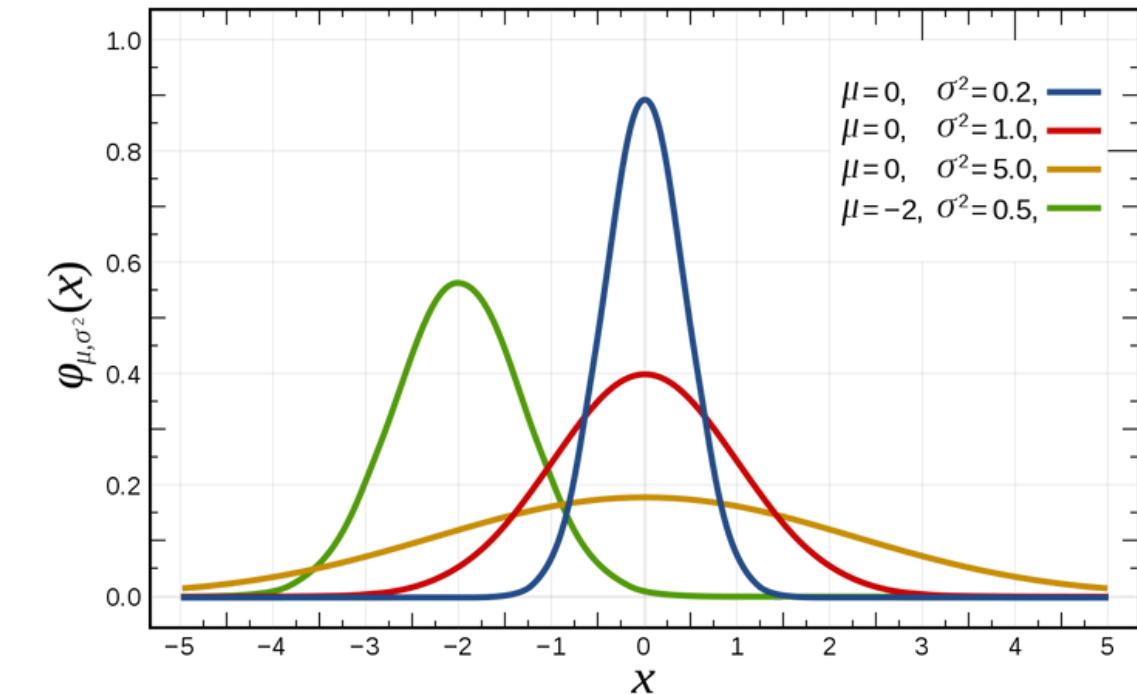
ກຸບລົງຂອງ Naive Bayes ເປັນຫຼື່ງໃນວິລກອຣິກິມໃນກາຮຈໍາແນກປະເກດຂໍ້ອມູລ (Classification) ກີ່ໃຊ້ກຸບລົງກາຮສົລືຕີແລະກາຮດຳນວນຄວາມນ່າຈະເປັນ (Probability) ເພື່ອກຳນາຍຮີ້ວຈໍາແນກຂໍ້ອມູລໃນກລຸ່ມຮີ້ວປະເກດຕ່າງ ຫຼື ໂດຍ Naive Bayes ໃຊ້ກຸບລົງຂອງ ກຸບລົງກາຮຄວາມນ່າຈະເປັນ (Bayesian Probability) ໃນກຳນາຍ



$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

Likelihood      Class Prior Probability  
Posterior Probability      Predictor Prior Probability

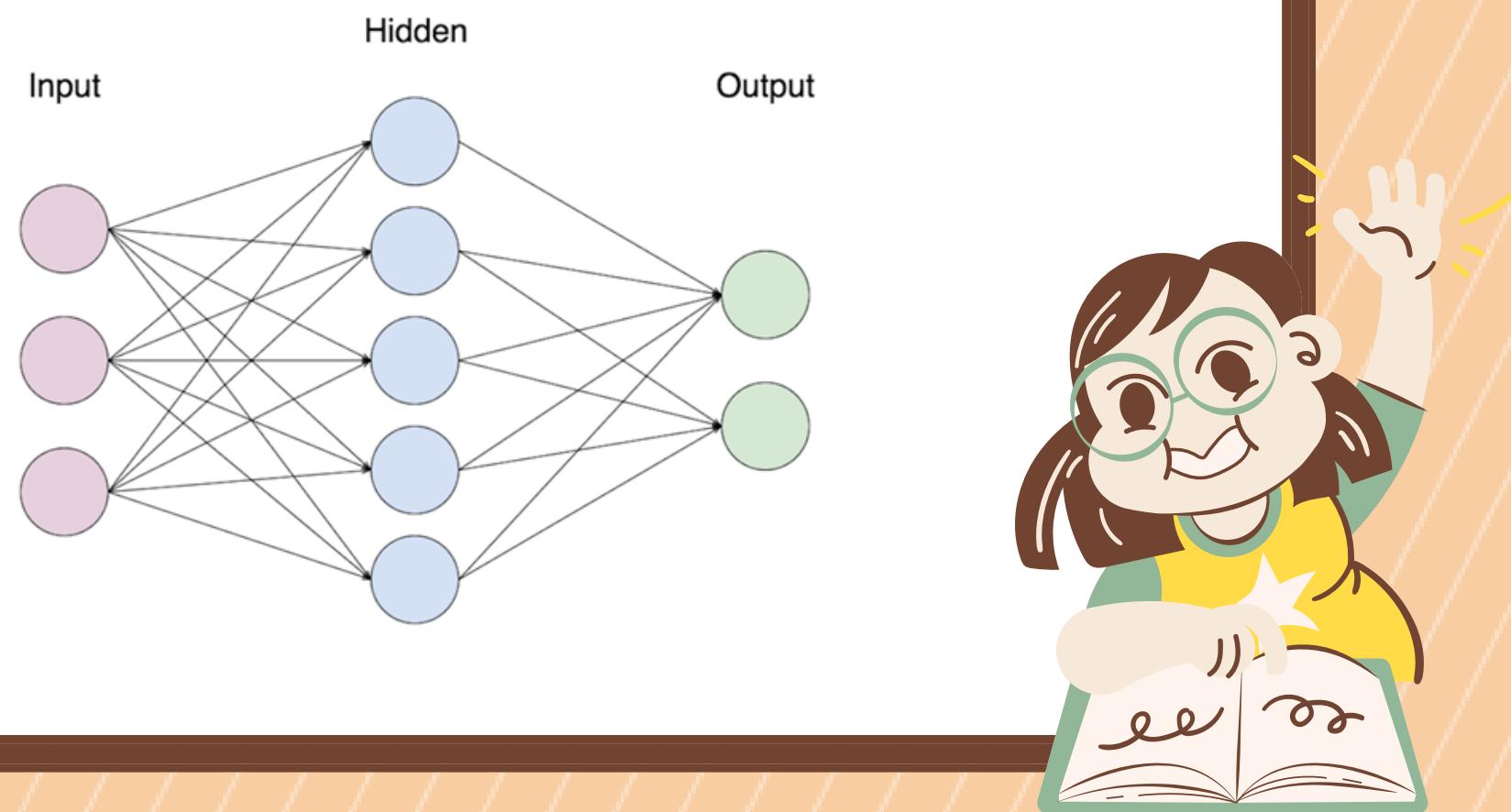
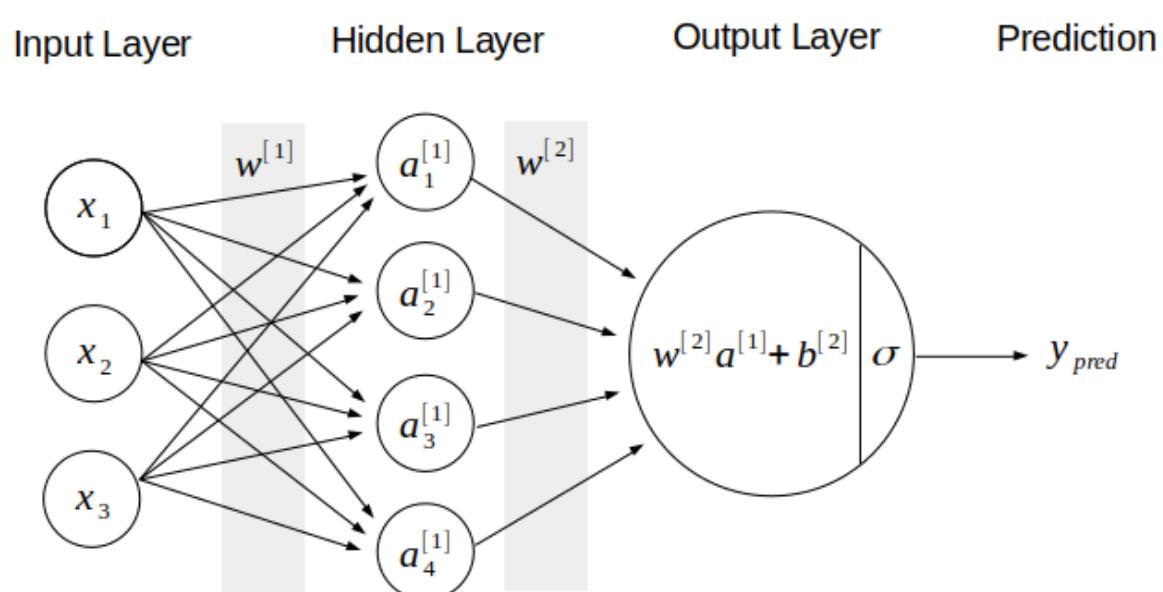
$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \cdots \times P(x_n | c) \times P(c)$$



# ☆ 🔒 ກຸບຖືຂອງ Neural Network Q

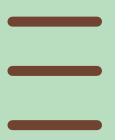


ກຸບຖືຂອງ Neural Network (ເຄຣືອຫ່າຍປະສາກ) ເປັນກະຮບວນກາຮກາງຄລິຕາສຕົມແລະ  
ຄວມພິວເຕອຮ໌ໃໝ່ເວັນໜີທີ່ຈໍາລວງກາຮກາງກຳຈານຂອງຮະບບປະສາກຂອງມຸນຸ່ຍ ໂດຍໃນກາງກຸບຖືນີ້  
ຖຸກອອກແບບໃໝ່ມີລັກສະນະຄລ້າຍກັບໂຄຮງສ້າງຂອງຮະບບປະສາກກາງຊົວກາພ ໂດຍມີໜ່ວຍ  
ປະມວລຜລເຮີຍກວ່າ "ໄຟດ" (neurons) ທີ່ກຳໜ້າທີ່ຮັບຂໍ້ມູນ (input), ປະມວລຜລ, ແລະສົ່ງ  
ຜລລັບຮົກໄປຢັ້ງໂຟດກັດໄປໜີອືຜລລັບຮົກທີ່ໃນກາງກຸບຖືເຮີຍກວ່າ "output layer"





# เนตுผลกីលើកមិនលេខ Neural Network Q



jakplukki opeapren mihel ghangmud 5 tawnin mii Neural Network kii mii c'a AUC sogn  
kii sogn doyle mii c'a ooyukii 0.963 kales Naive bayes ooyukii 0.956 tsing ikal deiyng kgnmaka reajing  
kamkaran tesdoyle izs xomnul 30% kii yung naym ideytes maa tes oslscrupwajak xomnul kii ide  
Naive bayes AUC ooyukii 0.956 hemion dem swn Neural Network mikiar pemjinnma  
0.966 tsing xomnul trang nii ppen netu plukki kama hra leik Neural Network

**Test and Score (1) - Orange**

**ផល Test and Score**

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.908	0.875	0.873	0.874	0.875	0.720
SVM (1)	0.498	0.604	0.519	0.493	0.604	-0.087
Logistic Regression (1)	0.937	0.866	0.867	0.869	0.866	0.709
Naive Bayes (1)	0.958	0.906	0.905	0.905	0.906	0.790
Neural Network (1)	0.963	0.916	0.915	0.916	0.916	0.812

Cross validation  
Number of folds: 5  
Stratified  
Cross validation by feature  
Random sampling  
Repeat train/test: 3

ផលលងការណាតម្លែងបង្កើតម្នល់

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.961	0.905	0.904	0.905	0.905	0.788
SVM (1)	0.654	0.623	0.541	0.537	0.623	-0.018
Logistic Regression (1)	0.937	0.859	0.861	0.864	0.859	0.697
Naive Bayes (1)	0.958	0.907	0.906	0.906	0.907	0.792
	0.966	0.918	0.917	0.918	0.918	0.816

Show performance scores Target class: (Average over classes)

31.9k | 31.9k | 31.9k | 5x31923

# การสอน ไมเดล



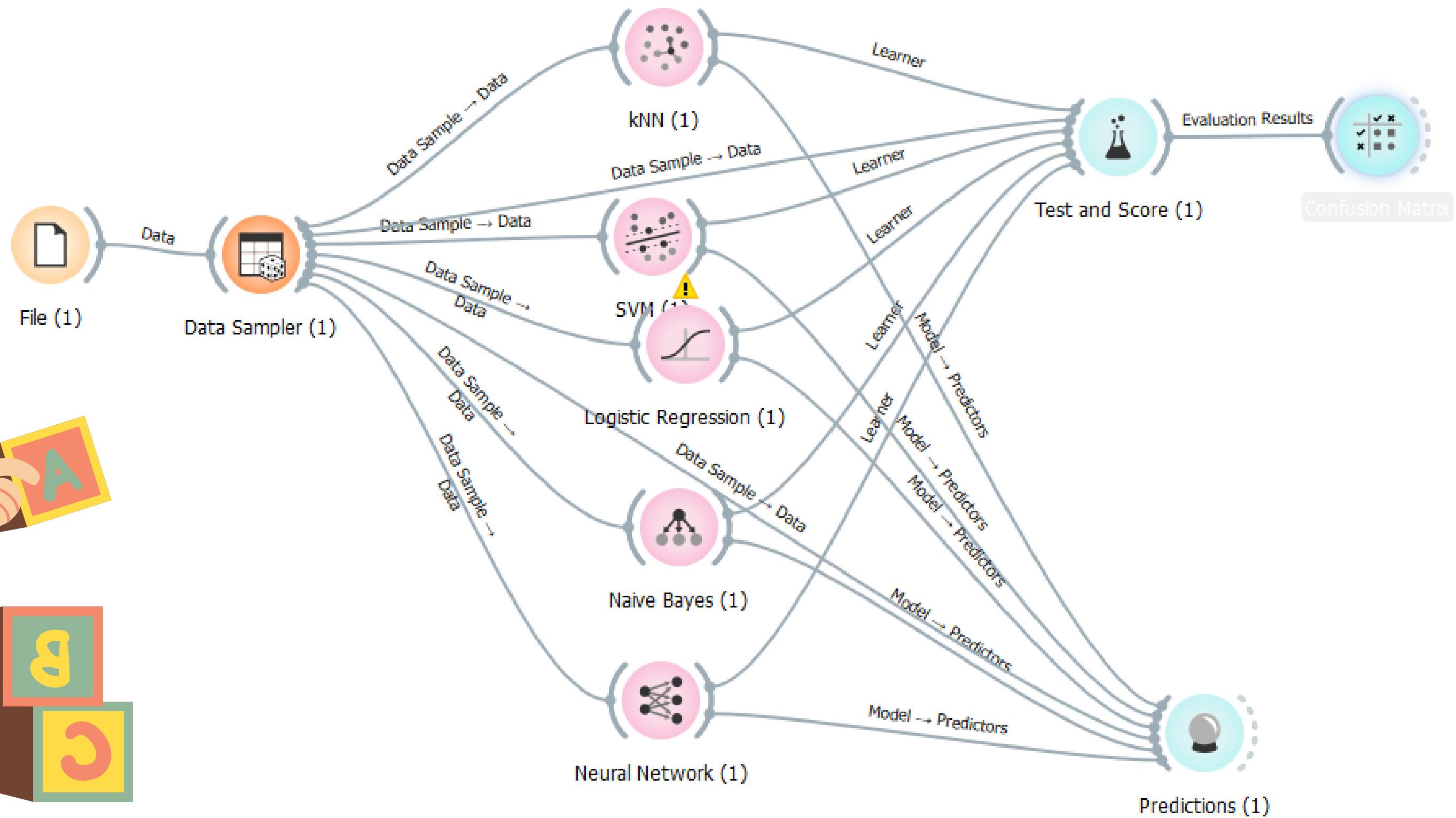
RULE 01



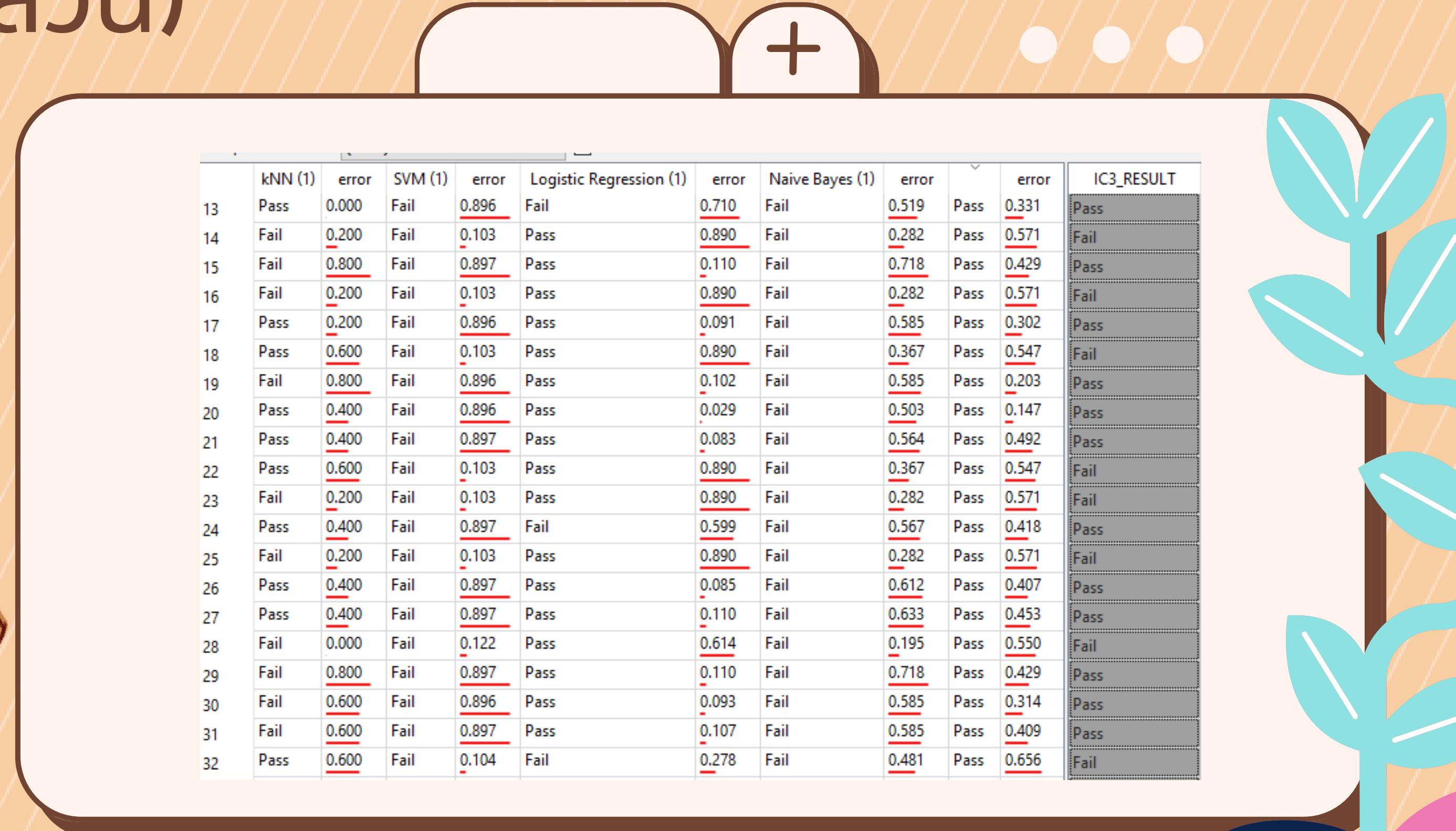
ใช้เครื่องมือ DATA SAMPLE ในการมาเทรน โดยแทนอยู่ที่ 70/30 โดยข้อมูล 70 % นั้นเอาไปเทรน และนำอีก 30 % ไป tes หลังจากนั้น ไปดูค่า TEST AND SCORE เพื่อดูค่าความแม่นยำของโมเดล ซึ่งกลุ่มพากเราต้องการค่าความเชื่อ มั่นอยู่ที่ 0.95 ขึ้นไปซึ่งมี 2 ตัวที่มีค่ามากกว่า 0.95 คือ Naive Bayes และ Neural Network ที่ค่าความเชื่อมั่นอยู่ที่ 0.956 และ 0.963 จากทั้งหมด 5 โมเดล และทำการเพิ่มลดข้อมูลว่ามีข้อมูลไหนที่มีผลมากอันไหนมีผลน้อยต่อ การสร้างโมเดล



# Tools and Model In Orange



# ผลลัพธ์การ คำนาย(บางส่วน)



	kNN (1)	error	SVM (1)	error	Logistic Regression (1)	error	Naive Bayes (1)	error	Pass	error	IC3_RESULT
13	Pass	0.000	Fail	0.896	Fail	0.710	Fail	0.519	Pass	0.331	Pass
14	Fail	0.200	Fail	0.103	Pass	0.890	Fail	0.282	Pass	0.571	Fail
15	Fail	0.800	Fail	0.897	Pass	0.110	Fail	0.718	Pass	0.429	Pass
16	Fail	0.200	Fail	0.103	Pass	0.890	Fail	0.282	Pass	0.571	Fail
17	Pass	0.200	Fail	0.896	Pass	0.091	Fail	0.585	Pass	0.302	Pass
18	Pass	0.600	Fail	0.103	Pass	0.890	Fail	0.367	Pass	0.547	Fail
19	Fail	0.800	Fail	0.896	Pass	0.102	Fail	0.585	Pass	0.203	Pass
20	Pass	0.400	Fail	0.896	Pass	0.029	Fail	0.503	Pass	0.147	Pass
21	Pass	0.400	Fail	0.897	Pass	0.083	Fail	0.564	Pass	0.492	Pass
22	Pass	0.600	Fail	0.103	Pass	0.890	Fail	0.367	Pass	0.547	Fail
23	Fail	0.200	Fail	0.103	Pass	0.890	Fail	0.282	Pass	0.571	Fail
24	Pass	0.400	Fail	0.897	Fail	0.599	Fail	0.567	Pass	0.418	Pass
25	Fail	0.200	Fail	0.103	Pass	0.890	Fail	0.282	Pass	0.571	Fail
26	Pass	0.400	Fail	0.897	Pass	0.085	Fail	0.612	Pass	0.407	Pass
27	Pass	0.400	Fail	0.897	Pass	0.110	Fail	0.633	Pass	0.453	Pass
28	Fail	0.000	Fail	0.122	Pass	0.614	Fail	0.195	Pass	0.550	Fail
29	Fail	0.800	Fail	0.897	Pass	0.110	Fail	0.718	Pass	0.429	Pass
30	Fail	0.600	Fail	0.896	Pass	0.093	Fail	0.585	Pass	0.314	Pass
31	Fail	0.600	Fail	0.897	Pass	0.107	Fail	0.585	Pass	0.409	Pass
32	Pass	0.600	Fail	0.104	Fail	0.278	Fail	0.481	Pass	0.656	Fail

# Q เปรียบเทียบและวิเคราะห์ผล



	kNN (1)	error	SVM (1)	error	Logistic Regression (1)	error	Naive Bayes (1)	error	error	IC3_RESULT	
13	Pass	0.000	Fail	<u>0.896</u>	Fail	<u>0.710</u>	Fail	<u>0.519</u>	Pass	<u>0.331</u>	Pass
14	Fail	<u>0.200</u>	Fail	<u>0.103</u>	Pass	<u>0.890</u>	Fail	<u>0.282</u>	Pass	<u>0.571</u>	Fail
15	Fail	<u>0.800</u>	Fail	<u>0.897</u>	Pass	<u>0.110</u>	Fail	<u>0.718</u>	Pass	<u>0.429</u>	Pass
16	Fail	<u>0.200</u>	Fail	<u>0.103</u>	Pass	<u>0.890</u>	Fail	<u>0.282</u>	Pass	<u>0.571</u>	Fail
17	Pass	<u>0.200</u>	Fail	<u>0.896</u>	Pass	<u>0.091</u>	Fail	<u>0.585</u>	Pass	<u>0.302</u>	Pass
18	Pass	<u>0.600</u>	Fail	<u>0.103</u>	Pass	<u>0.890</u>	Fail	<u>0.367</u>	Pass	<u>0.547</u>	Fail
19	Fail	<u>0.800</u>	Fail	<u>0.896</u>	Pass	<u>0.102</u>	Fail	<u>0.585</u>	Pass	<u>0.203</u>	Pass
20	Pass	<u>0.400</u>	Fail	<u>0.896</u>	Pass	<u>0.029</u>	Fail	<u>0.503</u>	Pass	<u>0.147</u>	Pass
21	Pass	<u>0.400</u>	Fail	<u>0.897</u>	Pass	<u>0.083</u>	Fail	<u>0.564</u>	Pass	<u>0.492</u>	Pass
22	Pass	<u>0.600</u>	Fail	<u>0.103</u>	Pass	<u>0.890</u>	Fail	<u>0.367</u>	Pass	<u>0.547</u>	Fail
23	Fail	<u>0.200</u>	Fail	<u>0.103</u>	Pass	<u>0.890</u>	Fail	<u>0.282</u>	Pass	<u>0.571</u>	Fail
24	Pass	<u>0.400</u>	Fail	<u>0.897</u>	Fail	<u>0.599</u>	Fail	<u>0.567</u>	Pass	<u>0.418</u>	Pass
25	Fail	<u>0.200</u>	Fail	<u>0.103</u>	Pass	<u>0.890</u>	Fail	<u>0.282</u>	Pass	<u>0.571</u>	Fail
26	Pass	<u>0.400</u>	Fail	<u>0.897</u>	Pass	<u>0.085</u>	Fail	<u>0.612</u>	Pass	<u>0.407</u>	Pass
27	Pass	<u>0.400</u>	Fail	<u>0.897</u>	Pass	<u>0.110</u>	Fail	<u>0.633</u>	Pass	<u>0.453</u>	Pass

จากการทำนายพบว่า มีข้อมูลบางตัวที่ SVM, KNN หรือ Logistic Regression มีการทำนายตรงกับ Naive Bayes และ Neural Network ในบางข้อมูล แต่ผลการทำนายโดยรวมแล้ว Neural Network และ Naive Bayes มีการทำนายแม่นยำกว่าโดยเฉลี่ย



# សម្រាប់



=



		Predicted		$\Sigma$
		Fail		
Actual	Fail	17249	1549	18798
	Pass	3046	6887	9933
$\Sigma$		20295	8436	28731

KNN

		Predicted		$\Sigma$
		Fail		
Actual	Fail	16918	1880	18798
	Pass	8736	1197	9933
$\Sigma$		25654	3077	28731

SVM

		Predicted		$\Sigma$
		Fail		
Actual	Fail	16466	2332	18798
	Pass	1578	8355	9933
$\Sigma$		18044	10687	28731

Logistic Regression

		Predicted		$\Sigma$
		Fail		
Actual	Fail	17701	1097	18798
	Pass	2648	7285	9933
$\Sigma$		20349	8382	28731

Naive Bayes

		Predicted		$\Sigma$
		Fail		
Actual	Fail	17713	1085	18798
	Pass	2262	7671	9933
$\Sigma$		19975	8756	28731

Neural Network

# สรุปผล



จากตาราง Confused Matrix นั้นค่าการทำนายได้ที่สุดคือ Neural Network จากโมเดลทั้งหมดที่ได้ทดลองใช้ ซึ่งเหตุผลที่มีความแม่นยำสูง เพราะ ข้อมูลที่เรา拿来ไปสร้างโมเดลนั้นมีเยอะและมีความหลากหลายมาก อีกทั้งยังมีตัวแปรบางตัวที่ส่งต่อการทำนายมาก เช่น โดยตัวแปร ONLINE\_ASSIGNMENT\_SUBMISSION\_FREQUENCY , IC3\_EXAM\_TIMEUSED ที่มีผลต่อการทำนายอย่างมาก ต่างจากตัวแปร ASS\_ACADEMIC\_YEAR, CLASS\_SEMESTER ที่มีการส่งผ่านน้อยมาก ในการทำโมเดล แต่ถ้าเอาออกไปก็ส่งผลต่อการทำนายไม่เหลือ

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.908	0.875	0.873	0.874	0.875	0.720
SVM (1)	0.498	0.604	0.519	0.493	0.604	-0.087
Logistic Regression (1)	0.937	0.866	0.867	0.869	0.866	0.709
Naive Bayes (1)	0.958	0.906	0.905	0.905	0.906	0.790
Neural Network (1)	0.963	0.916	0.915	0.916	0.916	0.812

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.880	0.840	0.837	0.838	0.840	0.638
SVM (1)	0.521	0.631	0.562	0.566	0.631	0.032
Logistic Regression (1)	0.936	0.864	0.865	0.867	0.864	0.706
Naive Bayes (1)	0.913	0.870	0.867	0.870	0.870	0.706
Neural Network (1)	0.938	0.884	0.882	0.883	0.884	0.738

# สรุปผล



ก่อนนำตัวแปรที่ชื่อว่า ASS\_ACADEMIC\_YEAR และ CLASS\_SEMESTER ออก

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.908	0.875	0.873	0.874	0.875	0.720
SVM (1)	0.498	0.604	0.519	0.493	0.604	-0.087
Logistic Regression (1)	0.937	0.866	0.867	0.869	0.866	0.709
Naive Bayes (1)	0.958	0.906	0.905	0.905	0.906	0.790
Neural Network (1)	0.963	0.916	0.915	0.916	0.916	0.812

หลังนำตัวแปรที่ชื่อว่า ASS\_ACADEMIC\_YEAR และ CLASS\_SEMESTER ออก

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.880	0.840	0.837	0.838	0.840	0.638
SVM (1)	0.521	0.631	0.562	0.566	0.631	0.032
Logistic Regression (1)	0.936	0.864	0.865	0.867	0.864	0.706
Naive Bayes (1)	0.913	0.870	0.867	0.870	0.870	0.706
Neural Network (1)	0.938	0.884	0.882	0.883	0.884	0.738



# Thank You

นาย ฤทธิฤณ เมืองขวัญ 116510907040-9

น.ส.นภัสสร ปานพร้อม 116510907015-1

