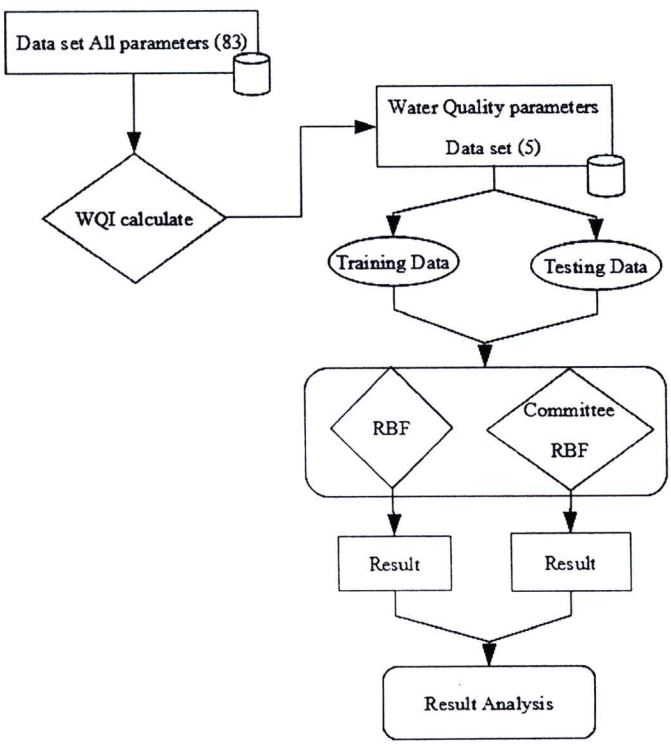
บทที่ 3
วิธีการดำเนินงานวิจัย

การดำเนินการวิจัยในวิทยานิพนธ์ฉบับนี้ เกี่ยวกับการวัดประสิทธิภาพของอัลกอริทึมที่ใช้ในการวิเคราะห์ข้อมูล ในส่วนของบทที่ 3 จะกล่าวถึงการดำเนินการทดลองซึ่งแบ่งเป็น 2 ส่วนที่สำคัญ คือ การวิเคราะห์ข้อมูลโดยโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และการวิเคราะห์ข้อมูลโดยกลไกเรเดียลเบสิสฟังก์ชันแบบคณะกรรมการ ซึ่งการออกแบบและการกำหนดค่าตัวแปรในการทดลองจะกล่าวถึงในเนื้อหาต่อไป

1. ขั้นตอนของการทดลอง(Flow chart)

ในขั้นตอนของการทดลอง เพื่อให้ทราบถึงกระบวนการที่ชัดเจน จึงมีการแสดงแผนภูมิการทดลอง ดังรูปภาพ



ภาพที่ 11 แบบจำลองการพยากรณ์ค่าโดยกลไกเรเดียลเบสิสฟังก์ชันแบบคณะกรรมการ

ข้อมูลที่ได้รับจากกรมควบคุมมลพิษเป็นข้อมูลดิบที่ยังไม่มีการตรวจสอบความสมบูรณ์ของข้อมูล และยังไม่มีการคำนวณค่า WQI ซึ่งเป็นพารามิเตอร์ที่ใช้ในการวิจัย ดังนี้ในแผนภูมิจึงมีกระบวนการคำนวณค่า WQI ก่อนเข้าสู่กระบวนการแบ่งข้อมูล เพื่อใช้ในการวิเคราะห์โดยโครงข่ายประสาทเทียม ซึ่งมีการวิเคราะห์ผล 2 ส่วนคือ วิเคราะห์โดยโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชันและวิเคราะห์โดยกลไกเรเดียลเบสิสฟังก์ชันแบบคณะกรรมการและนำผลการทดลองมาวิเคราะห์ผลและสรุป

2. เครื่องมือที่ใช้ในการทำวิจัย

เครื่องคอมพิวเตอร์

หน่วยประมวลผลกลาง (CPU)	AMD Athlon X2	1.91	GHz
หน่วยความจำของระบบ (RAM)	DDR2	2	GB
ฮาร์ดดิสก์ (Hard disk)	SATA	160	GB
การ์ดแสดงผล (VGA Card)	Geforce 9600GT	512	MB
แผงวงจรหลัก (Main Board)	Asrock SATA2 FSB	1	GHz

โปรแกรมคอมพิวเตอร์

Windows 7

Clementine 12.0

Matlab 2008

3. การเตรียมข้อมูล

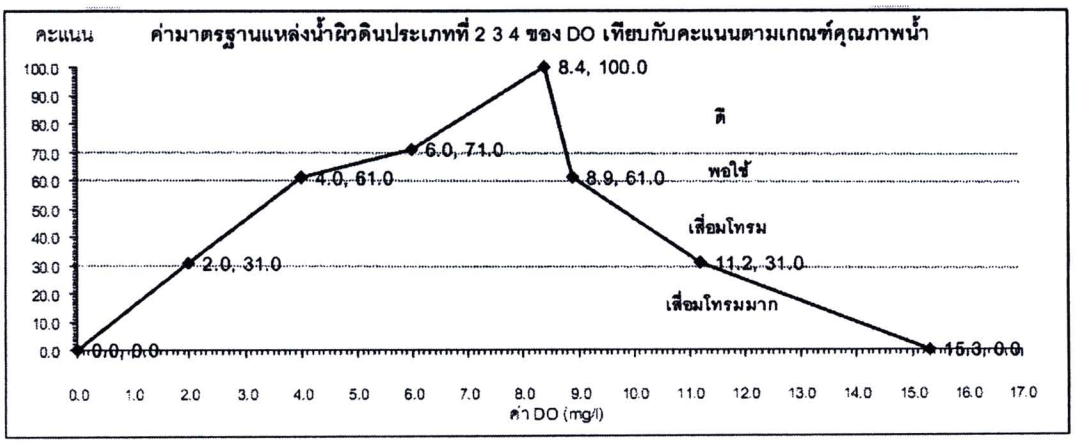
ปัจจุบันกรมควบคุมมลพิษ สำนักจัดการคุณภาพน้ำ ได้ทำการเก็บข้อมูลคุณภาพน้ำแหล่งน้ำจืดในลุ่มน้ำ 25 ลุ่มน้ำหลัก โดยแบ่งเป็นแม่น้ำ 48 แม่น้ำ 4 แหล่งน้ำนิ่ง และดูคลองต่าง ๆ มีจุดเก็บรวมทั้งสิ้น 553 จุด ผู้วิจัยได้เลือกข้อมูลของแหล่งน้ำจำนวน 11 แหล่งในภาคตะวันออกเฉียงเหนือ ซึ่งประกอบด้วย แม่น้ำพอง แม่น้ำชี แม่น้ำมูล แม่น้ำลำปาว แม่น้ำเสียว หนองหาน แม่น้ำสงคราม แม่น้ำเลย แม่น้ำอูน แม่น้ำลำชีและแม่ลำตะคอง ข้อมูลที่ได้ คือ พารามิเตอร์ที่แสดงคุณสมบัติทางกายภาพและชีวภาพของแหล่งน้ำ ซึ่งข้อมูลเหล่านี้สามารถนำมาคำนวณ เพื่อหาค่าเกณฑ์ของคุณภาพน้ำในแหล่งน้ำนั้นๆได้ และกรมควบคุมมลพิษได้ กำหนดวิธีการในการคำนวณค่า Water Quality Index (WQI) เพื่อใช้ในการกำหนดเกณฑ์คุณภาพน้ำ โดยวิธีการในการคำนวณเป็นดังนี้

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

3.1 การคำนวณค่า Water quality index (WQI)

การประเมินคุณภาพน้ำของแหล่งน้ำผิวดินโดยทั่วไป [7] ใช้พารามิเตอร์คุณภาพน้ำทั่วไป (WQI) ที่มีหน่วยเป็นคะแนน เริ่มจาก 0 ถึง 100 คะแนน 91–100 คะแนน ถือว่า คุณภาพน้ำอยู่ในเกณฑ์ดีมาก 71-90 คะแนน คุณภาพน้ำอยู่ในเกณฑ์ดี 61-70 คะแนน คุณภาพน้ำอยู่ในเกณฑ์พอใช้ 31-60 คะแนน คุณภาพน้ำอยู่ในเกณฑ์เสื่อมโทรม 0-30 คะแนนคุณภาพน้ำอยู่ในเกณฑ์เสื่อมโทรมมาก คะแนนได้จากการรวมคะแนน พารามิเตอร์คุณภาพน้ำ 5 พารามิเตอร์ ได้แก่ ออกซิเจนละลายน้ำ (DO), แบคทีเรียกลุ่มฟีคัลโคลิฟอร์ม (Fecal Coliform Bacteria , FCB), การปนเปื้อนของแบคทีเรียกลุ่มโคลิฟอร์มทั้งหมด (TCB) ,แอมโมเนีย (NH_3 -N) และความสกปรกในรูปสารอินทรีย์ (Biological Oxygen Demand , BOD) เข้าด้วยกันเป็นคะแนนรวม

3.1.1 ออกซิเจนละลายน้ำ (DO) ใช้ในการประเมินประเภทแหล่งน้ำผิวดิน สามารถบ่งชี้ถึงความเหมาะสมในการดำรงชีวิตของสัตว์น้ำทั่วไป โดยรวมของแหล่งน้ำมีปัจจัยหลายอย่าง ที่ทำให้มีค่ามากขึ้นหรือน้อยลง ทั้งนี้น้ำเสียจากแหล่งกำเนิดต่างๆก็เป็นปัจจัยหนึ่ง จากภาพที่ 12 แสดงให้เห็นถึงค่าคะแนนคุณภาพน้ำ เมื่อเปรียบเทียบกับ ค่า DO ซึ่งค่า DO ที่ให้คะแนนคุณภาพน้ำสูงสุด คือ 8.4



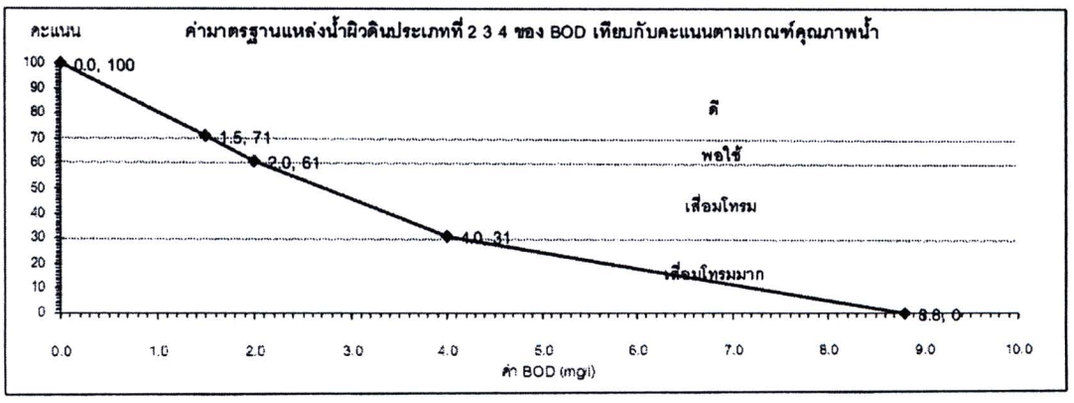
ภาพที่ 12 การเทียบคะแนนตามเกณฑ์คุณภาพน้ำกับค่ามาตรฐานแหล่งน้ำผิวดินประเภทที่ 2 3 4 ของ DO

การคิดค่าคะแนนตามเกณฑ์คุณภาพน้ำ สามารถคำนวณตามสมการในตารางที่ 1 โดยสมการที่ใช้ขึ้นอยู่กับค่า DO ที่จะคำนวณ คะแนนที่ได้อยู่ในช่วง 0 – 100 และจะนำคะแนนที่ได้ไปใช้ในการคำนวณ WQI ต่อไป

ตารางที่ 1 สมการการคิดคะแนนเทียบกับค่า DO

ค่า DO	สูตรสมการในการคิดคะแนน
0.0 – 4.0 mg/l	คะแนน = 15.25*(ค่า DO)+0.1667
4.1 – 6.0 mg/l	คะแนน = 5*(ค่า DO)+41
6.1 – 8.4 mg/l	คะแนน =12.083*(ค่าDO)-1.5
8.5 – 8.9 mg/l	คะแนน = -78*(ค่าDO)+755.2
9.0 – 11.2 mg/l	คะแนน = -13.043*(ค่าDO)+177.09
11.3 – (>=15.3) mg/l	คะแนน = -7.561*(ค่าDO)+115.68

3.1.2 ความสกปรกในรูปสารอินทรีย์ (Biological Oxygen Demand, BOD) ใช้ในการประเมินประเภทแหล่งน้ำผิวดิน สามารถบ่งชี้ถึง ความสกปรกของแหล่งน้ำ สาเหตุสำคัญคือน้ำเสียของแหล่งกำเนิดจากชุมชน อุตสาหกรรม และเกษตรกรรม จากภาพที่ 13 แสดงให้เห็นถึงค่าคะแนนคุณภาพน้ำ เมื่อเปรียบเทียบกับ ค่า BOD ซึ่งค่า BOD ที่ให้คะแนนคุณภาพน้ำสูงสุด คือ 0



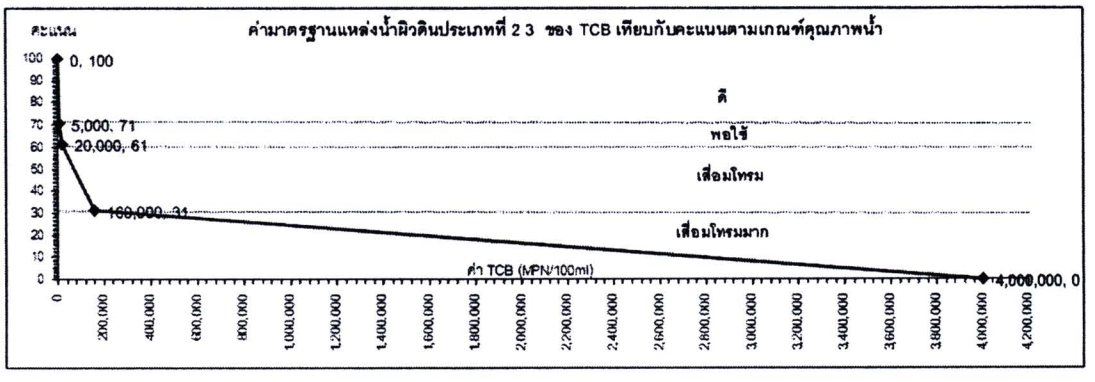
ภาพที่ 13 การเทียบคะแนนตามเกณฑ์คุณภาพน้ำ กับ ค่ามาตรฐานแหล่งน้ำผิวดินประเภทที่ 2 3 4 ของ BOD

การคิดค่าคะแนนตามเกณฑ์คุณภาพน้ำ สามารถคำนวณตามสมการในตารางที่ 2โดยสมการที่ใช้ขึ้นอยู่กับค่า BOD ที่จะคำนวณ คะแนนที่ได้อยู่ในช่วง 0 – 100 และจะนำคะแนนที่ได้ไปใช้ในการคำนวณ WQI ต่อไป

ตารางที่ 2 สมการการคิดคะแนนเทียบกับค่า BOD

ค่า BOD	สูตรสมการในการคิดคะแนน
0.0 – 1.5 mg/l	คะแนน = -19.333(ค่าBOD) + 100
1.6 – 2.0 mg/l	คะแนน = -20(ค่าBOD) + 101
2.1 – 4.0 mg/l	คะแนน = -15(ค่าBOD) + 91
4.1 – (>=8.8) mg/l	คะแนน = -6.4583*(ค่าBOD)+56.833

3.1.3 การปนเปื้อนของแบคทีเรียกลุ่มโคลิฟอร์มทั้งหมด (TCB) ใช้ในการประเมินประเภทแหล่งน้ำผิวดิน สามารถบ่งชี้ถึงการปนเปื้อนแบคทีเรียกลุ่มโคลิฟอร์มจากธรรมชาติโดยครอบคลุมถึงกลุ่มฟีคัลโคลิฟอร์ม จากสิ่งขับถ่ายในลำไส้ของสัตว์เลือดอุ่น ใช้วิเคราะห์ร่วมกับ FCB จากภาพที่ 14 แสดงให้เห็นถึงค่าคะแนนคุณภาพน้ำ เมื่อเปรียบเทียบกับ ค่า TCB ซึ่งค่า TCB ที่ให้คะแนนคุณภาพน้ำสูงสุด คือ 0



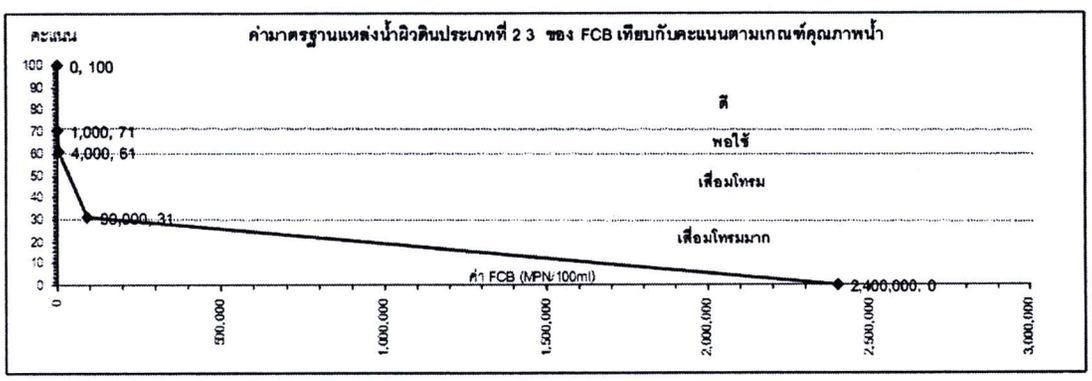
ภาพที่ 14 การเทียบคะแนนตามเกณฑ์คุณภาพน้ำ กับ ค่ามาตรฐานแหล่งน้ำผิวดินประเภทที่ 2 3 4 ของ TCB

การคิดค่าคะแนนตามเกณฑ์คุณภาพน้ำ สามารถคำนวณตามสมการในตารางที่ 3 โดยสมการที่ใช้ขึ้นอยู่กับค่า TCB ที่จะคำนวณ คะแนนที่ได้อยู่ในช่วง 0 – 100 และจะนำคะแนนที่ได้ไปใช้ในการคำนวณ WQI ต่อไป

ตารางที่ 3 สมการการคิดคะแนนเทียบกับค่า TCB

ค่า TCB(MPN/100ml)	สูตรสมการในการคิดคะแนน
0.0 – 5,000	คะแนน = -0.0058*(ค่าTCB)+100
5,001 – 20,000	คะแนน = -0.0007*(ค่าTCB) + 74.333
20,001 – 160,000	คะแนน = -0.0002*(ค่าTCB) + 65.286
>160,000	คะแนน = -8E-06*(ค่าTCB) + 32.292

3.1.4 แบคทีเรียกลุ่มฟีคัลโคลิฟอร์ม (Fecal Coliform Bacteria, FCB) ใช้ในการประเมินประเภทแหล่งน้ำผิวดินสามารถบ่งชี้ถึงการปนเปื้อนแบคทีเรียกลุ่มฟีคัลโคลิฟอร์ม จากสิ่งขับถ่ายในลำไส้ของสัตว์เลือดอุ่นที่สำคัญคือ คน และหมู สาเหตุสำคัญคือน้ำเสียจากชุมชน ฟาร์มหมู จากภาพที่ 15 แสดงให้เห็นถึงค่าคะแนนคุณภาพน้ำ เมื่อเปรียบเทียบกับ ค่า FCB ซึ่งค่า FCB ที่ให้คะแนนคุณภาพน้ำสูงสุด คือ 0



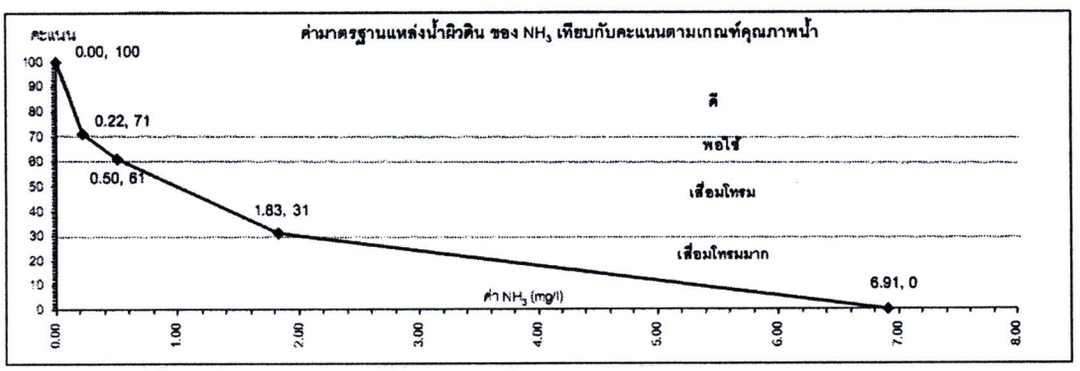
ภาพที่ 15 การเทียบคะแนนตามเกณฑ์คุณภาพน้ำ กับ ค่ามาตรฐานแหล่งน้ำผิวดินประเภทที่ 2 3 4 ของ FCB

การคิดค่าคะแนนตามเกณฑ์คุณภาพน้ำ สามารถคำนวณตามสมการในตารางที่ 4โดยสมการที่ใช้ขึ้นอยู่กับค่า FCB ที่จะคำนวณ คะแนนที่ได้อยู่ในช่วง 0 – 100 และจะนำคะแนนที่ได้ไปใช้ในการคำนวณ WQI ต่อไป

ตารางที่ 4 สมการการคิดคะแนนเทียบกับค่า FCB

ค่า FCB(MPN/100ml)	สูตรสมการในการคิดคะแนน
0.0 – 1,000	คะแนน = -0.029*(ค่าFCB) + 100
1,001 – 4,000	คะแนน = -0.0033*(ค่าFCB) + 74.333
4,001 – 90,000	คะแนน = -0.0003*(ค่าFCB) + 62.395
>90,000	คะแนน = -1E-05*(ค่าFCB) + 32.208

3.1.5 แอมโมเนีย (NH_3) สามารถบ่งชี้ถึงการปนเปื้อนน้ำเสียจากกิจกรรมมนุษย์ได้แก่ การขับถ่าย ปุ๋ยจากการเกษตร อาหารสัตว์น้ำที่เหลือตกค้าง จากภาพที่ 16 แสดงให้เห็นถึงค่าคะแนนคุณภาพน้ำ เมื่อเปรียบเทียบกับ ค่า NH_3 ซึ่งค่า NH_3 ที่ให้คะแนนคุณภาพน้ำสูงสุด คือ 0



ภาพที่ 16 การเทียบคะแนนตามเกณฑ์คุณภาพน้ำ กับ ค่ามาตรฐานแหล่งน้ำผิวดินประเภทที่ 2 3 4 ของ NH_3

การคิดค่าคะแนนตามเกณฑ์คุณภาพน้ำ สามารถคำนวณตามสมการในตารางที่ 5โดยสมการที่ใช้ขึ้นอยู่กับค่า NH_3 ที่จะคำนวณ คะแนนที่ได้อยู่ในช่วง 0 – 100 และจะนำคะแนนที่ได้ไปใช้ในการคำนวณ WQI ต่อไป

ตารางที่ 5 สมการการคิดคะแนนเทียบกับค่าแอมโมเนีย

ค่า NH_3 (mg/l)	สูตรสมการในการคิดคะแนน
0.0 – 0.22	คะแนน = -131.82*(ค่าNH_3) + 100
0.23 – 0.50	คะแนน = -35.714*(ค่าNH_3) + 78.857
0.51 – 1.83	คะแนน = -22.556*(ค่าNH_3) + 72.278
>1.83	คะแนน = -6.1024*(ค่าNH_3) + 42.167

เพื่อให้การคำนวณค่าพารามิเตอร์มีความแม่นยำและสอดคล้องกับคุณภาพน้ำผิวดิน จึงมีการพิจารณาแต่ละสถานีตรวจวัดค่าควรใช้วิธีการพิจารณาแต่ละจุดตรวจวัด ว่ามีคุณภาพน้ำโดยรวมอยู่ในเกณฑ์ ดี พอใช้ เสื่อมโทรม หรือเสื่อมโทรมมาก จากนั้นจึงเทียบกับ ค่าเฉลี่ยคะแนน 5 พารามิเตอร์ ถ้าตรงกัน คะแนนพิเศษ เท่ากับ 0 ถ้าต่างกัน 1 ระดับ คะแนนพิเศษเท่ากับ 10 ถ้าต่างกัน 2 ระดับ คะแนนพิเศษเท่ากับ 15 โดยมีเงื่อนไขดังนี้

ตารางที่ 6 เกณฑ์การกำหนดค่าคะแนนพิเศษ

พารามิเตอร์	ข้อกำหนดในการพิจารณาเกณฑ์คุณภาพน้ำ			
DO	≥ 4.0 mg/l	≥ 2.5 mg/l	≥ 2.0 mg/l	≥ 0.0 mg/l
	ดี	พอใช้	เสื่อมโทรม	เสื่อมโทรมมาก
BOD	≤ 1.5 mg/l	≤ 2.0 mg/l	≤ 4.0 mg/l	> 4.0 mg/l
	ดี	พอใช้	เสื่อมโทรม	เสื่อมโทรมมาก
TCB	≤ 5,000 หน่วย	≤ 20,000 หน่วย	> 20,000 หน่วย	
	ดี	พอใช้	เสื่อมโทรม	
FCB	≤ 1,000 หน่วย	≤ 4,000 หน่วย	> 4,000 หน่วย	
	ดี	พอใช้	เสื่อมโทรม	
NH_3	≤ 0.22 mg/l	≤ 0.50 mg/l	≤ 1.83 mg/l	> 1.83 mg/l
	ดี	พอใช้	เสื่อมโทรม	เสื่อมโทรมมาก

Pollution Control Department

หลังจากได้เกณฑ์คุณภาพน้ำแต่ละพารามิเตอร์แล้ว เมื่อจะสรุปว่า จุดตรวจวัดนั้นอยู่ในเกณฑ์ใดก็จะพิจารณาจากเกณฑ์คุณภาพน้ำที่ต่ำที่สุด จากนั้นเทียบกับค่าคะแนนเฉลี่ย 5 พารามิเตอร์โดยค่าคะแนนเทียบกับเกณฑ์คุณภาพน้ำเป็น ดังนี้

ค่าคะแนนคุณภาพน้ำ มากกว่าหรือเท่ากับ 71 เกณฑ์คุณภาพน้ำ คือ ดี

ค่าคะแนนคุณภาพน้ำ มากกว่าหรือเท่ากับ 61 แต่น้อยกว่า 70 เกณฑ์คุณภาพน้ำ คือ พอใช้

ค่าคะแนนคุณภาพน้ำ มากกว่าหรือเท่ากับ 31 แต่น้อยกว่า 60 เกณฑ์คุณภาพน้ำ คือ เสื่อมโทรม

ค่าคะแนนคุณภาพน้ำ มากกว่าหรือเท่ากับ 0 แต่น้อยกว่า 30 เกณฑ์คุณภาพน้ำคือเสื่อมโทรมมาก

ถ้าเกณฑ์คุณภาพน้ำไม่ต่างกัน คะแนนพิเศษ คือ 0 ถ้าต่างกัน 1 ระดับ คะแนนพิเศษ คือ 10 ต่างกัน 2 ระดับ คะแนนพิเศษ คือ 15 ต่างกัน 3 ระดับ คะแนนพิเศษ คือ 20 นำคะแนนพิเศษไปลบกับ ค่าคะแนนเฉลี่ย 5 พารามิเตอร์ได้คะแนนรวมของจุดตรวจวัดนั้น สูตร คะแนนรวม = ค่าเฉลี่ยของคะแนนทั้ง 5 พารามิเตอร์ – คะแนนพิเศษ

ตารางที่ 7 ตัวอย่างการคำนวณค่า WQI

พารามิเตอร์	DO	BOD	TCB	FCB	NH_3
ค่าคุณภาพน้ำ	3.8	0.7	160,000	17,000	0.20
คะแนน	58	86	31	57	74
คะแนนเฉลี่ย	61				
เกณฑ์คุณภาพน้ำ	พอใช้				

พารามิเตอร์	DO	BOD	TCB	FCB	NH_3
ค่าคุณภาพน้ำ	3.8	0.7	160,000	17,000	0.20
เกณฑ์คุณภาพน้ำ	พอใช้	ดี	เสื่อมโทรม	เสื่อมโทรม	ดี
สรุปเกณฑ์คุณภาพน้ำ	เสื่อมโทรม				
พบว่าต่างกัน 1 ระดับ	10				
ค่าคะแนนรวม	61 – 10 = 51				

4. ชุดข้อมูลและคุณลักษณะข้อมูล

การทำความสะอาดข้อมูล

ข้อมูลที่ได้จากกรมควบคุมมลพิษ ยังไม่สามารถใช้ในการวิเคราะห์โดยโครงข่ายประสาทเทียมได้เนื่องจากข้อมูลได้แยกเป็นปีที่ทำการตรวจวัดและแยกตามสถานีตรวจวัด และประกอบด้วยพารามิเตอร์หลายชนิดที่ไม่มีส่วนเกี่ยวข้องในการวิเคราะห์ข้อมูล อีกทั้งยังประกอบด้วยสัญลักษณ์ทางคณิตศาสตร์และสัญลักษณ์พิเศษที่ต้องทำการปรับแก้เพื่อให้ข้อมูลทั้งหมดเป็นจำนวนจริง ดังนั้นจึงมีกระบวนการในการทำความสะอาดข้อมูล และ สิ่งผิดปกติที่พบในชุดข้อมูลมีดังนี้

ข้อมูลบางพารามิเตอร์มีไม่ครบ

สัญลักษณ์พิเศษในข้อมูล เช่น *

เครื่องหมายทางคณิตศาสตร์ เช่น < = >

หากข้อมูลที่ใช้ในการคำนวณ WQI ซึ่งต้องใช้ทั้งหมด 5 พารามิเตอร์ ใน 1 ตัวอย่าง ขาดหายไปเพียง พารามิเตอร์เดียวจะไม่สามารถคำนวณค่า WQI ในตัวอย่าง นั้นๆได้

รายละเอียดคุณลักษณะ

DO ออกซิเจนละลายน้ำ ใช้หน่วยวัดคือ มิลลิกรัมต่อลิตร และในชุดข้อมูลมีค่าอยู่ในช่วง 0.4 - 13.3

BOD ความสกปรกในรูปสารอินทรีย์ ใช้หน่วยวัดคือ มิลลิกรัมต่อลิตร และในชุดข้อมูลมีค่าอยู่ในช่วง 0 - 13.6

TCB การปนเปื้อนของแบคทีเรียกลุ่มโคลิฟอร์มทั้งหมด ใช้หน่วยวัดคือ เอ็ม.พี.เอ็น ต่อ100 มิลลิลิตร และในชุดข้อมูลมีค่าอยู่ในช่วง 1 – 16,000,000

FCB แบคทีเรียกลุ่มฟีคัลโคลิฟอร์ม ใช้หน่วยวัดคือ เอ็ม.พี.เอ็น ต่อ100 มิลลิลิตร และในชุดข้อมูลมีค่าอยู่ในช่วง 1 – 5,000,000

NH_3 แอมโมเนีย ใช้หน่วยวัดคือ มิลลิกรัมต่อลิตร และในชุดข้อมูลมีค่าอยู่ในช่วง 0 – 30.46

Class เกณฑ์คุณภาพน้ำที่ได้จากการคำนวณค่า WQI แบ่งตามระดับดังนี้ 1 คือ เสื่อมโทรมมาก 2 คือ เสื่อมโทรม 3 คือ พอใช้ และ 4 คือ ดี

ตารางที่ 8 คุณลักษณะของชุดข้อมูล

Data Set Characteristics:	Multivariate
Attribute Characteristics:	Real
Associated Tasks:	Classification
Number of Instances:	1344
Number of Attributes:	6

ตัวอย่างชุดข้อมูลคุณภาพน้ำที่สามารถใช้ในการวิเคราะห์ข้อมูลโดยโครงข่ายประสาทเทียม

ตารางที่ 9 ตัวอย่างชุดข้อมูลที่ใช้ในงานวิจัย

DO	BOD	TCB	FCB	NH3	Class
5.17	2.27	400	2	0.19	3
4.8	1.3	2	2	0.14	4
6.2	0.98	230	130	0.01	4
6.6	0.8	400	200	0.11	4
6.9	0.9	400	2	0.02	4

การแบ่งชุดข้อมูล

ชุดข้อมูลที่ใช้ในการทดลองถูกแบ่งเป็น 2 ส่วนคือ

-ข้อมูลที่ใช้ในการเรียนรู้ของโครงข่าย(Training data)

-ข้อมูลที่ใช้ตรวจสอบความถูกต้องของโครงข่าย (Testing data)

การแบ่งข้อมูลผู้วิจัยได้กำหนดไว้ 3 รูปแบบ คืออัตราส่วนระหว่าง training data และ testing data เป็น 70:30 , 60:40 และ 50:50 โดยการกำหนดข้อมูลเป็น 3 แบบ เพื่อเป็นเกณฑ์ในการวัดศักยภาพในการเรียนรู้ของโครงข่าย ในกรณีที่ข้อมูลในการเรียนรู้มีไม่เท่ากัน

ตารางที่ 10 การแบ่งข้อมูลเพื่อเรียนรู้และตรวจสอบความถูกต้องของแบบจำลอง

Training Data		Testing Data	
Records	Percent	Records	Percent
672	50	672	50
806	60	538	40
941	70	403	30

5. การเตรียมโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมเป็นเทคนิคที่จะนำมาใช้ในการวิเคราะห์ข้อมูล ดังนั้นจึงมีการกำหนดค่าที่เกี่ยวข้องเพื่อให้ สามารถใช้ประสิทธิภาพของโครงข่ายได้อย่างเต็มประสิทธิภาพ

5.1 โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน เป็นโครงข่ายประสาทเทียมที่มีการทำงานแบบไปข้างหน้า ประกอบด้วย ชั้น Input ชั้น Hidden และ ชั้น Output ในการวิเคราะห์ข้อมูลนั้นต้องมีการปรับค่าพารามิเตอร์ เพื่อให้ได้โครงข่ายที่มีประสิทธิภาพในการคัดแยกข้อมูลมากที่สุด โดยใช้เวลาในการคำนวณไม่มากนัก ซึ่งพารามิเตอร์ที่ต้องปรับค่ามีดังนี้

RBF clusters จำนวน Activate Function หรือ Cluster ที่ใช้ในโครงข่ายซึ่งค่าที่กำหนดมีผล ต่อจำนวนนิวรอลในชั้น Hidden layer

Persistence จำนวนรอบในการเรียนรู้ ที่โครงข่ายจะทำการเรียนรู้ไปเรื่อยๆ หากไม่มีการเปลี่ยนแปลงเกิดขึ้น

Eta ค่าอัตราการเรียนรู้ของโครงข่ายประสาทเทียม หากใช้โครงข่ายแบบเรเดียลเบสิสฟังก์ชั่น ในแต่ละรอบการเรียนรู้ค่านี้จะไม่มีการเปลี่ยนแปลง ถ้าหากใช้ โครงข่ายประสาทเทียมแบบอื่นๆ เช่น MLP ค่า Eta จะมีการปรับค่าในแต่ละรอบการเรียนรู้ กำหนดค่าได้ 2 แบบ คือ

แบบสุ่ม จะมีการปรับเปลี่ยนค่าขึ้นกับรอบการเรียนรู้ 2 รอบแรก

แบบค่าคงที่ ผู้วิจัยจะต้องกำหนดค่าเองให้เหมาะสมกับการใช้งานและค่านี้จะไม่มีการเปลี่ยนแปลงในขณะที่มีกระบวนการเรียนรู้ของโครงข่าย

Alpha ค่าโมเมนตัมที่ใช้ในการปรับค่าน้ำหนัก ในขณะกระบวนการเรียนรู้ กำหนดค่าได้ในช่วง 0 - 1

RBF overlapping ในโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชันมีการใช้งาน Cluster ในการทำงานซึ่งมีขอบเขตของแต่ละ Clusters ค่าพารามิเตอร์นี้เป็นตัวกำหนดว่าขอบเขตของแต่ละ Clusters มีการซ้อนทับกันเท่าไร โดยปกติในการเรียนรู้ตัวอย่างจะมีผลต่อ Clusters ที่มีความคล้ายกันมากที่สุด การเพิ่มค่าพารามิเตอร์นี้ส่งผลให้ระเบียนข้อมูลได้รับผลกระทบจาก Clusters ที่อยู่ระยะห่างไกลมากขึ้น

ตารางที่ 11 พารามิเตอร์ที่ปรับค่าในโครงข่ายประสาทเทียมเรเดียลเบสิสฟังก์ชั่น

Parameters	Values	Lower Bound	Upper Bound
Eta	0.5 , 0.7 , 0.9 ,1.2 ,1.5	0	10
Training Data	50% , 60% ,70%	0%	100%
Alpha	0.2	0	1
Clusters	800 , 900 , 1000	0	ไม่จำกัด

ตัวแปรสำคัญที่ใช้ในการเรียนรู้ของโครงข่าย คือ อัตราการเรียนรู้ (Learning Rate) ในงานวิจัยนี้คือ ค่า Eta ซึ่งสามารถปรับค่าได้ในช่วง 0 – 10 ในการทดลองเบื้องต้นพบว่าค่าที่เหมาะสมอยู่ในช่วง 0.5 – 1.5 จึงมีการกำหนดค่าในช่วงนี้ในการทดลองดังในตารางที่ โดยค่าที่มีผลต่อความแม่นยำมากคือช่วง 0.5 – 0.9 จึงมีการแบ่งเพิ่มค่าทีละ 0.2 เพื่อให้เห็นผลของความแม่นยำที่มีการเปลี่ยนแปลงในการปรับค่าในช่วงเหล่านี้ และจำนวนข้อมูลที่ใช้ในการเรียนรู้ของโครงข่าย (Training data) สามารถให้โครงข่ายเรียนรู้ข้อมูลได้สูงสุด 100 % แต่ในการวิจัยทางด้านโครงข่ายประสาทเทียมค่าที่นิยมตามหลักสากลคือ 70% และผู้วิจัยแบ่งข้อมูลเป็น 3 แบบ คือ 50%, 60%,70% เพื่อต้องการทราบว่าจำนวนข้อมูลที่ใช้เรียนรู้ต่างกันจะให้ผลความแม่นยำที่ต่างกันหรือไม่ และการให้แบบจำลองเรียนรู้ข้อมูลที่มากกว่า 70% อาจทำให้เกิดปัญหาโครงข่ายจดจำข้อมูลแต่ไม่มีการเรียนรู้ข้อมูลได้ และจำนวนฟังก์ชันกระตุ้นที่ใช้ (Clusters) การกำหนดค่านี้ได้มีการทดลองเบื้องต้นและพบว่าค่าที่เหมาะสมอยู่ในช่วง 800 – 1000 Clusters จึงมีการกำหนดค่าดังตารางที่ และจากการทดลองพบว่าเมื่อเพิ่มจำนวนมากเกินกว่าค่าที่กำหนดไว้ไม่ทำให้ความแม่นยำของโครงข่ายเพิ่มมากขึ้น และค่า Alpha สามารถกำหนดค่าได้ในช่วง 0 – 1 ในการทดลองเบื้องต้นพบว่าการปรับค่านี้ไม่มีผลต่อความแม่นยำมากนัก จึงกำหนดค่าไว้ที่ 0.2

สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน มีทั้งหมด 3 ชั้น

ชั้นแรกคือชั้น Input ในงานวิจัยจะมีทั้งหมด 5 โหนด คือ จำนวนพารามิเตอร์ 5 ค่าที่ใช้ในการวิเคราะห์ค่าคุณภาพน้ำ

ชั้นที่สอง คือ ชั้น Hidden เป็นชั้นที่ใช้ในการวิเคราะห์ข้อมูลประกอบด้วย Activate function จำนวนนิวรอลในชั้นนี้ขึ้นอยู่กับค่า Clusters ที่ใช้ ซึ่งให้ประสิทธิภาพที่แตกต่างกันในแต่ละ

ค่าที่กำหนด โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชั่นจะมีชั้น Hidden เพียงแค่ชั้นเดียวจึงมีผลให้การเรียนรู้ทำได้เร็วกว่า โครงข่ายประสาทเทียมแบบอื่นๆที่ สามารถมีชั้น Hiden ได้หลายๆชั้น Activate function ที่ใช้ในชั้นนี้ คือ Gaussian function

ชั้นสุดท้ายคือ ชั้น Output เป็นชั้นแสดงผลข้อมูลที่ได้จากการวิเคราะห์ผล ในงานวิจัยนี้จะมีทั้งหมด 4 นิวรอล ซึ่งก็คือจำนวน class ของคุณภาพน้ำ

การเรียนรู้ของโครงข่าย

กระบวนการเรียนรู้ของโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน มีหลายวิธีซึ่งแต่ละวิธีการจะให้ผลกำหนดจุดศูนย์กลางของฟังก์ชันที่ต่างกัน วิธีการที่ใช้ในงานวิจัยนี้คือ

Fixed Center Selected at Random

วิธีการนี้จุดศูนย์กลางของฟังก์ชัน จะถูกเลือกแบบสุ่มจากข้อมูลที่ใช้ในการเรียนรู้ และ Activate function ในชั้น Hidden ก็คือ เรเดียลเบสิสฟังก์ชั่นที่ถูกกำหนดค่าไว้ตายตัว ซึ่งฟังก์ชันที่ใช้ คือ Gaussian function ที่ค่าเบี่ยงเบนมาตรฐานถูกกำหนดโดยการกระจายตัวของจุดศูนย์กลางของฟังก์ชันดังสมการ

$$G(\|\mathbf{x}-\mathbf{t}_i\|) = \exp(-\frac{M}{d^2}\|\mathbf{x}-\mathbf{t}_i\|^2), \qquad i = 1, 2, ...,M \tag{6}$$

โดย M คือ จำนวนของจุดศูนย์กลางของฟังก์ชัน และ d คือระยะทางที่มากที่สุดระหว่างจุดศูนย์กลางที่กำลังพิจารณา ในกระบวนการคัดแยกข้อมูลที่ใช้ Gaussian function เป็นเบสิสฟังก์ชันจะต้องมีการกำหนดพารามิเตอร์การกระจาย σ โดยปกติแล้วจะกำหนดดังสมการ

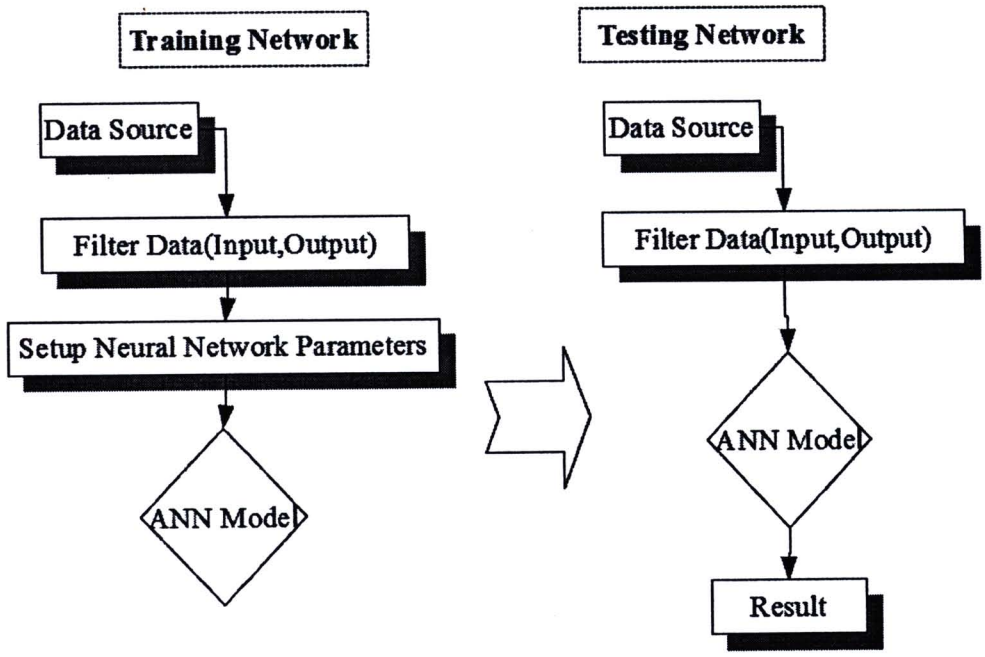
$$\sigma = \frac{d}{\sqrt{2M}} \tag{7}$$

การเรียนรู้วิธีนี้พารามิเตอร์ที่สำคัญคือ ค่าน้ำหนักของของชั้น Output ดังนั้นการฝึกสอนโครงข่ายที่ง่ายที่สุด คือกำหนดให้เวคเตอร์จุดศูนย์กลางมีค่าคงที่ และวิธีการนี้สิ่งสำคัญอย่างหนึ่ง คือ จำนวนของจุดศูนย์กลางของฟังก์ชันที่สุ่มเลือกมานั้นจะต้องมีปริมาณมากพอที่จะคลอบคลุม Input

ในกระบวนการคัดแยกข้อมูลของโครงข่ายแบบเรเดียลเบสิสฟังก์ชั่น มีการทำงานแบ่งเป็น 2 ส่วน คือกระบวนการในการเรียนรู้ของโครงข่าย เพื่อสร้างแบบจำลองที่มีความสามารถในการคัดแยกข้อมูลได้ดังรูป ได้แสดงกระบวนการ Training Network ซึ่งประกอบด้วยการนำเข้าชุดข้อมูลที่ใช้ใน

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

งานวิจัย และทำการกำหนดพารามิเตอร์ในส่วน Input และ Output ที่ต้องการวิเคราะห์ และทำการกำหนดค่าพารามิเตอร์ของโครงข่าย ซึ่งได้กล่าวรายละเอียดแล้วในขั้นต้นและทำการเรียนรู้ เพื่อสร้างแบบจำลอง ในส่วนของกระบวนการ Testing network เมื่อได้แบบจำลองที่มีความสามารถในการคัดแยกข้อมูลได้แล้ว จึงมีการนำเข้าชุดข้อมูลเพื่อทดสอบความแม่นยำและวิเคราะห์ผลที่ได้จากแบบจำลอง

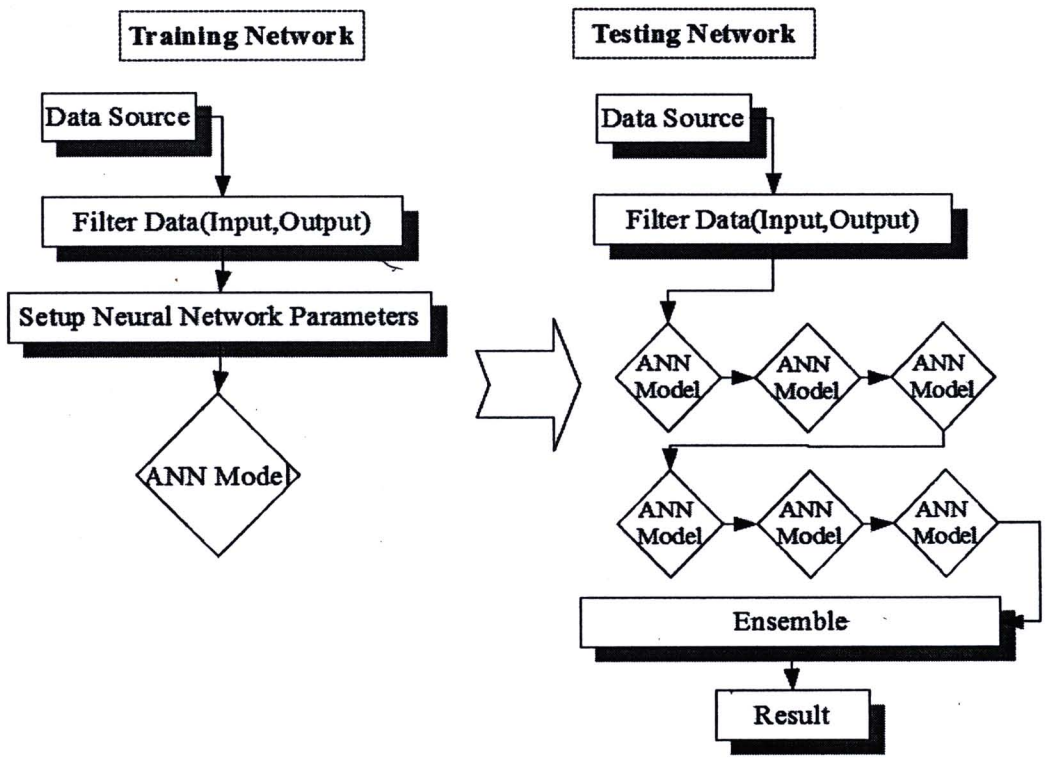


ภาพที่ 17 การคัดแยกข้อมูลโดยโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน

5.2 กลไกเรเดียลเบสิสฟังก์ชันแบบคณะกรรมการ

กลไกแบบคณะกรรมการ ถูกใช้งานอย่างแพร่หลายในการวิเคราะห์ข้อมูล โดยนักวิจัยต้องการที่จะเพิ่มประสิทธิภาพของโครงข่ายประสาทเทียมให้สูงขึ้น โดยหลักการง่ายๆ คือ การให้แบบจำลองโครงข่ายประสาทเทียมหลายๆตัว ทำงานร่วมกันและทำการรวมผลลัพธ์ของแต่ละแบบจำลอง

Ensemble Node คือ การรวมผลของโครงข่าย จำนวน 2 โครงข่ายขึ้นไปเพื่อให้ได้ผลการพยากรณ์ข้อมูลที่แม่นยำมากขึ้นกว่าโครงข่ายเพียงตัวเดียวและการรวมผลการพยากรณ์จากหลายๆ โครงข่ายจะช่วยลดข้อจำกัดของ การพยากรณ์เพียงโครงข่ายอันเดียวอีกด้วย ในการคัดแยกข้อมูลการรวมผลของการพยากรณ์จะใช้วิธีการ Voting



ภาพที่ 18 การคัดแยกข้อมูลโดยกลไกเรเดียลเบสิสฟังก์ชันแบบคณะกรรมการ

กลไกแบบคณะกรรมการที่ใช้ในงานวิจัย จะประกอบด้วยโครงข่ายแบบเรเดียลเบสิสฟังก์ชั่นหลายๆโครงข่ายช่วยกันทำงาน ดังภาพที่ ในกระบวนการ Training Network จะมีการสร้างแบบจำลองที่วิเคราะห์ข้อมูลขึ้นหลายตัวและนำแบบจำลองที่ได้ไปใช้ในกระบวนการ Testing Network และผลลัพธ์ของการคัดแยกข้อมูลในแบบจำลองแต่ละตัวจะถูกรวมกันโดยกระบวนการ Ensemble เพื่อให้ได้ผลลัพธ์สุดท้าย

6. การวัดประสิทธิภาพของการทดลอง

การวัดประสิทธิภาพของโครงข่ายประสาทเทียมได้จากค่าความถูกต้องของการคัดแยกข้อมูลและเปรียบเทียบกับข้อมูลทั้งหมด

$$\%\text{ความแม่นยำ} = \frac{\text{จำนวนตัวอย่างที่คัดแยกข้อมูลได้ถูกต้องจากแบบจำลอง} \times 100}{\text{จำนวนตัวอย่างข้อมูลทั้งหมด}} \quad (8)$$

ในบทที่ 3 ได้แสดงวิธีการในการเตรียมข้อมูล โดยเลือกใช้เฉพาะพารามิเตอร์คุณภาพน้ำที่เกี่ยวข้องกับเกณฑ์คุณภาพน้ำและกระบวนการในการทำความสะอาดข้อมูล เพื่อให้ได้ชุดข้อมูลที่สมบูรณ์ และได้คำนวณค่า WQI เพื่อใช้เป็นส่วนหนึ่งในชุดข้อมูล ที่ใช้งานในการวิเคราะห์ข้อมูล และได้แสดงถึงวิธีการที่จะดำเนินการวิจัยโดยได้แบ่งเป็น 2 ส่วนหลักๆ คือ การวิเคราะห์ข้อมูลโดยโครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชัน และการวิเคราะห์ข้อมูลโดยกลไกเรเดียลเบสิสฟังก์ชันแบบคณะกรรมการ ซึ่งในส่วนแรกนั้น ได้มีการแสดงถึงค่าพารามิเตอร์ที่ต้องกำหนดเพื่อให้ได้โครงข่ายประสาทเทียมแบบเรเดียลเบสิสฟังก์ชันที่ดีที่สุด และส่วนที่สองได้แสดงวิธีการที่จะทำการรวมผลลัพธ์ของการวิเคราะห์ข้อมูล ซึ่งแนวทางการทดลองที่ได้วางไว้และกระบวนการต่างๆที่ได้ตระเตรียมและวิธีการวัดผลที่ได้นำเสนอในบทนี้ จะเป็นแนวทางในการดำเนินการทดลองในบทที่ 4 ต่อไป