

# การคัดเลือกตัวแปรโดยการหาค่าเอยู่ซีเหมาะสมที่สุด

ปวิศ ธาริชาญ, วรรณ วรษ์เสรี

ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

## บทคัดย่อ

เอยู่ซีเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวจำแนก การวิเคราะห์ทางสถิติหาความสัมพันธ์ระหว่างเอยู่ซีและอัตราผิดพลาด (Error rate) พบว่ากรณีที่คลาสไม่สมดุลที่มีอัตราผิดพลาดสูง ตัวจำแนกที่มีความแม่นยำ (Accuracy) สูงอาจจะไม่ได้อัตราผิดพลาดสูง เนื่องจากความแม่นยำจะแปรผันตามจำนวนตัวอย่างที่จำแนกผิดพลาด ในขณะที่เอยู่ซีนอกจากจะแปรผันตามจำนวนตัวอย่างที่จำแนกผิดพลาดแล้วยังแปรผันตามลำดับ (Rank) ของตัวอย่างที่จำแนกผิดพลาดด้วย งานวิจัยนี้นำเสนอการทดลองเปรียบเทียบความสอดคล้องและความสามารถในการจำแนกกระหว่างดัชนีค่าความแม่นยำและค่า AUC พบว่า AUC มีความสอดคล้องกับความแม่นยำ อย่างไรก็ตาม AUC มีความสามารถในการจำแนกสูงกว่าความแม่นยำทั้งในกรณีที่คลาสสมดุลและไม่สมดุล นอกจากนี้ตัวจำแนกที่มีค่าเอยู่ซีสูงมีแนวโน้มที่จะมีความแม่นยำสูงด้วย การออกแบบตัวจำแนกที่มีค่าเอยู่ซีสูงจึงมีความเหมาะสมมากกว่าตัวจำแนกที่มีความแม่นยำสูง ดังนั้นจึงมีความจำเป็นในการหาฟังก์ชันความสูญเสีย (Loss function) ที่เหมาะสมสำหรับการหาค่าเอยู่ซีที่เหมาะสมที่สุด

**คำสำคัญ :** ความแม่นยำ เอยู่ซี เส้นโค้งอาร์โอซี ฟังก์ชันความสูญเสีย

## 1 บทนำ

เป้าหมายของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกคือการสร้างตัวจำแนกจากชุดข้อมูลที่บีบอัดเพื่อให้แบบจำลองสามารถใช้ ในการพยากรณ์ชุดข้อมูลทดสอบ โดยทั่วไปความสามารถในการทำนายของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกสามารถวัดได้จากค่าความแม่นยำ (หรือ อัตราผิดพลาดซึ่งเท่ากับ 1 ลบด้วยค่าความแม่นยำ) ของชุดข้อมูลทดสอบ และโดยส่วนใหญ่ของแบบจำลองการจำแนกนั้นสามารถประมาณค่าความน่าจะเป็นของการเกิดคลาส ได้แต่ยังไม่ค่อยนำมาประเมินประสิทธิภาพของแบบจำลองทำให้ความแม่นยำถูกพิจารณาเพียงถูกต้องหรือผิดพลาดเพียงเท่านั้น ค่าความแม่นยำนั้นอาจไม่เพียงพอในการประเมินประสิทธิภาพของแบบจำลองการจำแนก เช่น ในทางการแพทย์ที่ต้องการกระตุ้นยอดขายสูงสุดให้เพิ่มขึ้นจากลูกค้า จึงทำให้ต้องการดำเนินกลยุทธ์ทางการค้าต่อลูกค้าที่ส่งผลมากที่สุดต่อการขายในแต่ละบุคคล ไม่ใช่พิจารณาเฉพาะการดำเนินกลยุทธ์เพื่อให้ลูกค้าซื้อหรือไม่เท่านั้น ต้องการที่จะเพิ่มโอกาสการซื้อของลูกค้าให้เกิดผลสูงสุด ดังนั้นในกรณีนี้เพียงแค่เพิ่มโอกาสการซื้อของลูกค้านั้นไม่เพียงพอ แต่ต้องเป็นวิธีที่เพิ่มโอกาสการซื้อของลูกค้าได้มากที่สุดด้วย [1]

ดังนั้น การจัดอันดับ จึงเป็นที่ต้องการมากกว่าแค่การจัดประเภท และสามารถคำนวณได้ง่ายเนื่องจากแบบจำลองการจำแนกส่วนใหญ่จะสร้างการประมาณความน่าจะเป็นที่สามารถใช้ในการจัดอันดับได้

เส้นโค้ง ROC (Receiver Operating Characteristics) นำมาประยุกต์ใช้ในการประเมินประสิทธิภาพ [2, 3] การจัดอันดับของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนก [4, 5] โดยพบว่า AUC มีคุณสมบัติที่พึงประสงค์หลายประการเมื่อเทียบกับความแม่นยำ [6] ในบทความนี้จะทดลองเปรียบเทียบเพื่อพิสูจน์ว่า AUC เป็นการประเมินประสิทธิภาพแบบจำลองที่ดีกว่าความแม่นยำ

## 2 เกณฑ์สำหรับการเปรียบเทียบมาตรการการประเมิน

เริ่มต้นด้วยการเปรียบเทียบ AUC และความแม่นยำจากนั้นอธิบายคำจำกัดความที่เป็นทางการในการเปรียบเทียบตัวประเมินประสิทธิภาพแบบจำลองการจำแนกทั้งสองประเภท

### 2.1 ค่า AUC เทียบกับ ค่าความแม่นยำ

การคำนวณ AUC สามารถคำนวณได้จาก [7]

$$AUC = \frac{\sum r_i - n_p \frac{n_p + 1}{2}}{n_p n_n} \quad (1)$$

ตารางที่ 1 ตัวอย่างข้อมูลการคำนวณ AUC

	-	-	-	-	+	-	+	+	+	+
i					1		2	3	4	5
$r_i$					5		7	8	9	10

เมื่อ  $n_p$  และ  $n_n$  คือจำนวนตัวอย่างทั้งหมดของคลาสบวก และคลาสลบตามลำดับ และ  $r_i$  คือหมายเลขอันดับของคลาสบวกที่ i จากตัวอย่างในตารางที่ 1 พบว่ามีคลาสบวกและ คลาสลบอยู่อย่างละ 5 ตัว และเมื่อคำนวณค่า AUC จะได้ดังนี้  $\frac{(5+7+8+9+10)-5 \times \frac{6}{2}}{5 \times 5}$  ซึ่งเท่ากับ  $\frac{24}{25}$  โดยค่าสูงสุดของ AUC จะมีค่าเท่ากับ 1 ในตัวอย่างถัดไปจะเห็นว่าเหตุใด AUC จึงเป็นหน่วยวัดที่ดีกว่าความแม่นยำ

ตารางที่ 2 ตัวอย่าง แบบ จำลอง การ จำแนก ทั้ง สอง ที่มี ค่า ความ แม่นยำ เท่ากันแต่มี AUC ต่างกัน

ตัวจำแนกที่ 1	-	-	-	-	+	-	+	+	+	+
ตัวจำแนกที่ 2	+	-	-	-	-	+	+	+	+	-

พิจารณาตารางที่ 2 แบบจำลองการจำแนก 2 แบบจำลองมีการประมาณความน่าจะเป็นสำหรับชุดตัวอย่างการทดสอบ 10 ชุด โดยเป็นคลาสบวกและคลาสลบอย่างละ 5 ตัว พบว่าตัวจำแนกทั้งสอง มีค่าความแม่นยำเท่ากับ 80% (จำแนกคลาสบวกถูกต้อง(จริงบวก) 4 ตัว จำแนกคลาสลบถูกต้อง(จริงลบ) 4 ตัวและ จำแนกคลาสบวกผิด(เท็จลบ) 1 ตัว จำแนกคลาสลบผิด(เท็จบวก) 1 ตัว รวมถูกต้องทั้งหมด 8 ตัวจาก 10 ตัว) แต่ค่า AUC ของตัวจำแนกที่ 1 และ 2 นั้นเท่ากับ  $\frac{24}{25}$  และ  $\frac{16}{25}$  ตามลำดับ พบว่าความแม่นยำไม่สามารถแยกความแตกต่างของทั้งสองแบบจำลองได้ ในขณะที่ค่า AUC สามารถแยกความแตกต่างของทั้งสองแบบจำลองได้

ตาราง ที่ 3 ตัวอย่าง คำนวณ กรณีที่ ตัว จำแนก แรก มี ค่า AUC สูง กว่า ตัว จำแนกที่สองแต่ มีค่าความแม่นยำต่ำกว่าตัวจำแนกที่สอง

ตัวจำแนกที่ 3	-	-	-	+	+	-	-	+	+	+
ตัวจำแนกที่ 4	+	-	-	-	-	+	+	+	+	-

จากตารางที่ 3 พบว่าค่า AUC และค่าความแม่นยำสามารถได้ผลที่ขัดแย้งกันได้เพราะ ตัวจำแนกที่ 3 มีค่า AUC เท่ากับ  $\frac{21}{25}$  มากกว่า ตัวจำแนกที่ 4 ที่มีค่า AUC เท่ากับ  $\frac{16}{25}$  แต่ค่าความแม่นยำในตัวจำแนกที่ 3 เท่ากับ 60% น้อยกว่าตัวจำแนกที่ 4 ที่มีค่าความแม่นยำ 80% และก็ยังคงมีตัวอย่างคำนวณจากตารางที่ 4 ทั้งสองตัวจำแนกนั้นมีค่า AUC เท่ากัน แต่มีค่าความแม่นยำไม่เท่ากัน โดยตัวจำแนกที่ 5 และตัวจำแนกที่ 6 มีค่า AUC เท่ากันที่  $\frac{3}{5}$  แต่มีค่าความแม่นยำต่างกัน ตัวจำแนกที่ 5 มีค่าความแม่นยำที่ 60% และ ตัวจำแนกที่ 6 มีค่าความแม่นยำที่ 40%

ตาราง ที่ 4 ตัวอย่าง คำนวณ กรณีที่ ตัว จำแนก ทั้ง สอง มี ค่า AUC เท่า กัน แต่ มีค่าความแม่นยำแตกต่างกัน

ตัวจำแนกที่ 5	-	-	+	+	-	+	+	-	-	+
ตัวจำแนกที่ 6	-	-	+	+	+	-	-	+	-	+

ดังนั้นจึงต้องมีดัชนีสำหรับประเมินและเปรียบเทียบ ค่าความแม่นยำและ AUC ซึ่งในงานวิจัยนี้เสนอ 2 ดัชนีคือ ความสอดคล้องและความสามารถในการจำแนก

## 2.2 ความสอดคล้อง (Consistency) และความสามารถในการจำแนก (Discriminancy)

เมื่อต้องเปรียบเทียบประสิทธิภาพของแบบจำลองที่แตกต่างกันดัชนีหนึ่งที่เราควรคำนึงถึงคือความสอดคล้อง เพื่อระบุว่าแบบจำลองที่เปรียบเทียบกันนั้นจะมีการทำงานหรือ เปลี่ยนแปลงไปในทิศทางเดียวกัน หรือกลับกัน หรือไม่

กำหนดให้  $\Psi$  เป็นโดเมนทั้งหมดของลำดับแรงค์ที่เป็นไปได้  $AUC(a)$  เป็นค่า AUC ของลำดับชุดข้อมูล  $a$  และ  $acc(a)$  เป็นค่าความแม่นยำของชุดข้อมูล  $a$

**นิยามที่ 1** (ความสอดคล้อง). สำหรับสองการประเมินประสิทธิภาพ  $f, g$  ในโดเมน  $\Psi$  จะสอดคล้องกัน ถ้าไม่มี  $a, b \in \Psi$  ใดๆ ที่ทำให้  $f(a) < f(b)$  และ  $g(a) > g(b)$

อีกดัชนีหนึ่งความสามารถในการจำแนก ความสามารถในการแยกแยะรูปแบบที่แตกต่างกัน ดัชนีในการ ประเมิน ประสิทธิภาพแบบใด สามารถจำแนกสูงกว่าอีกแบบนั้นจะต้องมีเหตุการณ์ที่การประเมินประสิทธิภาพหนึ่งไม่สามารถ

แยกแยะรูปแบบสองชุดข้อมูลที่มีความต่างกันได้อีกการประเมินประสิทธิภาพสามารถทำได้

**นิยามที่ 2** (ความสามารถในการจำแนก). สำหรับ สอง การ ประเมิน ประสิทธิภาพ  $f, g$  ในโดเมน  $\Psi$  แล้ว  $f$  มีความสามารถในการจำแนกมากกว่า  $g$  ก็ต่อเมื่อถ้ากำหนดให้  $a, b \in \Psi$  แล้ว  $f(a) > f(b)$  และ  $g(a) = g(b)$  และต้องไม่มี  $a, b \in \Psi$  ใดๆ ที่ทำให้  $f(a) = f(b)$  และ  $g(a) > g(b)$

## 3 ทดลองเปรียบเทียบ

การทดลองเพื่อเปรียบเทียบประสิทธิภาพของดัชนีความแม่นยำและ AUC จะทดลองด้วยข้อมูลที่สามารถเกิดขึ้นในทุกกรณีโดยวิธีการเรียงสับเปลี่ยนทางคณิตศาสตร์ โดยข้อมูลจะมีทั้งหมดสองคลาสกำหนดให้เป็นคลาสบวกและ คลาสลบการทดลองจะแบ่งเป็นสองกรณีคือ ข้อมูลที่สมดุลกันและ ไม่สมดุลกัน

### 3.1 ข้อมูลสองคลาสที่สมดุล

ดังนั้นชุดข้อมูล ในการ ทดลอง นี้ จะ ประกอบ ด้วย ตัวอย่าง บวก และ ลบ จำนวนเท่ากัน (binary class) โดยจะทดลองข้อมูลที่มีขนาด 4, 6, 8, 10, 12, 14, 16, 18 และ 20 ตัวอย่างเมื่อข้อมูลมีขนาด  $2n$  จะมีรูปแบบที่เป็นไปได้ทั้งหมด  $\binom{2n}{n}$

ความหมายของความสอดคล้องและ ความสามารถในการจำแนกเป็นไปตามคำจำกัดความดังนี้

จากนิยามที่ 1 หมายถึงถ้ากำหนดให้  $a, b$  เป็นรูปแบบชุดข้อมูลที่แตกต่างกันแล้วถ้า  $AUC(a) > AUC(b)$  และ  $acc(a) > acc(b)$  ด้วย จะนับว่าค่า AUC มีความสอดคล้องกับค่าความแม่นยำ แต่ถ้า  $AUC(a) > AUC(b)$  และ  $acc(a) < acc(b)$  จะถูกนับว่าค่า AUC ไม่มีความสอดคล้องกับค่าความแม่นยำ นำไปสู่การคำนวณค่าระดับของความสอดคล้อง

**นิยามที่ 3** (ระดับของความสอดคล้อง). สำหรับสองการประเมินประสิทธิภาพ  $f, g$  ในโดเมน  $\Psi$  ให้  $R = \{(a, b) | a, b \in \Psi, f(a) > f(b), g(a) > g(b)\}$ ,  $S = \{(a, b) | a, b \in \Psi, f(a) > f(b), g(a) < g(b)\}$  ระดับของความสอดคล้อง  $C$  ของ  $f$  และ  $g$  โดยที่  $C(0 \leq C \leq 1)$  เมื่อ  $C = \frac{|R|}{|R|+|S|}$

จากนิยามที่ 3 หมายถึงระดับของความสอดคล้อง  $C$  คิดได้จากเหตุการณ์ที่ค่า AUC และค่าความแม่นยำ เป็นไปในทิศทางเดียวกัน ทารด้วย ผลรวมของทั้งสองเหตุการณ์ ผลลัพธ์จากการทดลองแสดงดังตารางที่ 5

จากนิยามที่ 2 หมายถึงถ้ากำหนดให้  $a, b$  เป็นรูปแบบชุดข้อมูลที่แตกต่างกันแล้วถ้า  $acc(a) = acc(b)$  แต่  $AUC(a) > AUC(b)$  ด้วย จะนับว่าค่า AUC นั้นมีความสามารถในการจำแนกสูงกว่าค่าความแม่นยำ แต่ เมื่อ  $AUC(a) = AUC(b)$  และ  $acc(a) > acc(b)$  จะถูกนับว่าค่าความแม่นยำ นั้นมีความสามารถในการจำแนกสูงกว่าค่า AUC นำไปสู่การคำนวณค่าระดับของความสามารถในการจำแนก

**นิยามที่ 4** (ระดับของความสามารถในการจำแนก). สำหรับสองการประเมิน ประสิทธิภาพ  $f, g$  ในโดเมน  $\Psi$  ให้  $P = \{(a, b) | a, b \in \Psi, f(a) > f(b), g(a) = g(b)\}$ ,  $Q = \{(a, b) | a, b \in \Psi, g(a) > g(b), f(a) = f(b)\}$  ระดับของความสามารถ

ในการจำแนก  $D$  ของ  $f$  ที่มีความสามารถในการจำแนกสูงกว่า  $g$  จะได้  $D = \frac{|P|}{|Q|}$

จากนิยามที่ 4 หมายถึงระดับของความสามารถในการจำแนก  $D$  คิดได้จากเหตุการณ์ที่ AUC มีความสามารถในการจำแนกสูงกว่าค่าความแม่นยำ หากด้วย เหตุการณ์ที่ค่าความแม่นยำ มีความสามารถในการจำแนกสูงกว่า AUC ผลลัพธ์จากการทดลองแสดงดังตารางที่ 6

**นิยามที่ 5.** ตัวประเมินประสิทธิภาพ  $f$  เป็นตัวประเมินประสิทธิภาพที่ดีกว่า  $g$  ก็ต่อเมื่อตัวประเมินประสิทธิภาพ  $f$  มีความสอดคล้องกับตัวประเมินประสิทธิภาพ  $g$  ก็ต่อเมื่อ  $C > 0.5$  และ ตัวประเมินประสิทธิภาพ  $f$  มีความสามารถในการจำแนกดีกว่า  $g$  ก็ต่อเมื่อ  $D > 1$

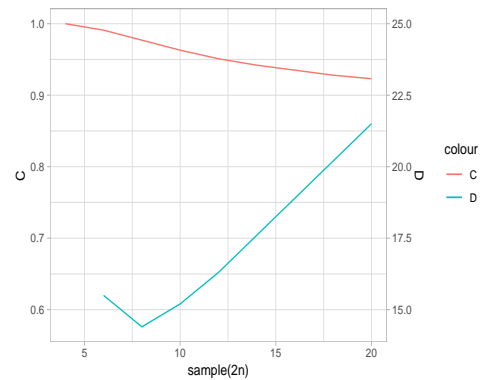
สังเกตในตารางที่ 5 และ 6 พบว่าในกรณีที่ข้อมูลมีความ สมดุลกัน ความสามารถในการจำแนกของค่า AUC สูงขึ้นเรื่อยๆ เมื่อจำนวนตัวอย่างมากขึ้นเรื่อยๆ และความสอดคล้องของตัว ประเมิน ประสิทธิภาพทั้งสองสูงมาก หมายความว่ากรณีที่ชุดตัวอย่างที่ สมดุล กัน ค่า AUC เป็น ตัว ประเมิน ประสิทธิภาพที่ดีกว่าค่าความแม่นยำ

ตารางที่ 5 ความสอดคล้องกันของ AUC และ ความแม่นยำ

$2n$	ค่า AUC และ ค่าความแม่นยำ สอดคล้องกัน	ค่า AUC และ ค่าความแม่นยำ ไม่สอดคล้องกัน	$C$
4	9	0	1
6	113	1	0.991
8	1,459	34	0.977
10	19,742	766	0.963
12	273,600	13,997	0.951
14	3,864,673	237,303	0.942
16	55,370,122	3,868,959	0.935
18	802,343,521	61,797,523	0.928
20	11,733,729,456	975,464,160	0.923

ตารางที่ 6 ความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ

$2n$	ค่า AUC จำแนกได้ดีกว่า ค่าความแม่นยำ	ค่าความแม่นยำ จำแนกได้ดีกว่า ค่า AUC	$D$
4	5	0	$\infty$
6	62	4	15.5
8	762	52	14.4
10	9,416	618	15.2
12	120,374	7,369	16.3
14	1,578,566	89,828	17.6
16	21,161,143	1,121,120	18.9
18	288,745,778	14,290,466	20.2
20	3,998,425,154	185,536,518	21.5



รูปที่ 1 ค่าความสอดคล้องและค่าความสามารถในการจำแนก เทียบจำนวนตัวอย่าง

ในการทดลองที่ข้อมูลเป็นมีจำนวนคลาสสองคลาสและเป็นข้อมูลที่สมดุลพบว่า AUC นั้นมีความสอดคล้องกับ ค่าความแม่นยำและ AUC นั้นสามารถจำแนกเหตุการณ์ที่แตกต่างกันที่ค่าความแม่นยำไม่สามารถจำแนกได้มากกว่า และเมื่อพิจารณาความสามารถในการจำแนกนั้นพบว่ายิ่งจำนวนข้อมูลเยอะมากขึ้นนั้น ความสามารถในการจำแนกของ AUC จะสูงขึ้น แสดงดังรูปที่ 1

### 3.2 ข้อมูลสองคลาสที่ไม่สมดุล

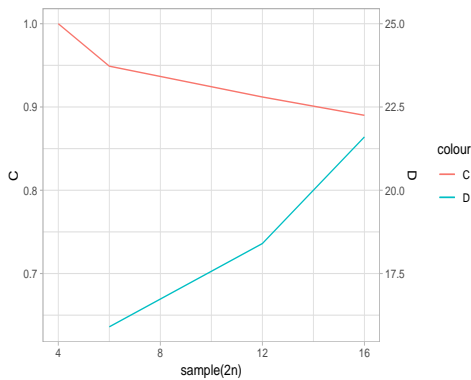
ข้อมูลไบนารีคลาสที่ไม่สมดุลโดยจะกำหนดให้มีตัวอย่างคลาสบวก 25% และตัวอย่างคลาสลบ 75% โดยข้อมูลที่ใช้จะมีจำนวน 4, 8, 12 และ 16 ตัวอย่าง และยังคงใช้สูตรการคำนวณค่า AUC เหมือนเดิมและเนื่องจากจำนวนตัวอย่างนั้นไม่สมดุลทำให้การคำนวณค่าความแม่นยำจะเปลี่ยนจากเดิมที่ให้ 5 ตัวอย่างแรกเป็นคลาสลบและ 5 ตัวอย่างถัดไปเป็นคลาสบวกหรืออีกนัยหนึ่งคือแบ่งตรงกลางอย่างละครึ่ง แต่เมื่อข้อมูลนั้นมีขนาดไม่เท่ากัน ทำให้การแยกคลาสบวกและคลาสลบเป็น 75% แรกเป็นคลาสลบ และ 25% เป็นต่อมาเป็นคลาสบวกตามอัตราส่วนของข้อมูลเข้าที่เปลี่ยนไป

ตารางที่ 7 ความสอดคล้องกันของ AUC และ ความแม่นยำ(ไม่สมดุล)

$2n$	ค่า AUC และ ค่าความแม่นยำ สอดคล้องกัน	ค่า AUC และ ค่าความแม่นยำ ไม่สอดคล้องกัน	$C$
4	3	0	1
6	187	10	0.949
12	12,716	1,225	0.912
16	926,884	114,074	0.890

ตารางที่ 8 ความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ(ไม่สมดุล)

$2n$	ค่า AUC จำแนกได้ดีกว่า ค่าความแม่นยำ	ค่าความแม่นยำ จำแนกได้ดีกว่า ค่า AUC	$D$
4	3	0	$\infty$
8	159	10	15.9
12	8,986	489	18.4
16	559,751	25,969	21.6



รูปที่ 2 ค่าความสอดคล้องและ ค่าความสามารถในการจำแนก เทียบจำนวน ตัวอย่าง กรณีข้อมูลไม่สมดุล

สังเกตในตารางที่ 7 และ 8 พบว่าในกรณีที่ข้อมูลมีความไม่สมดุลกัน ความสามารถในการจำแนกของค่า AUC สูงขึ้นเรื่อยๆ เมื่อจำนวนตัวอย่างมากขึ้นเรื่อยๆ และความสอดคล้องของตัว ประเมิน ประสิทธิภาพทั้งสองสูงมาก หมายความว่ากรณีชุดตัวอย่างที่ไม่สมดุลกันแสดงในรูปที่ 2

ค่า AUC เป็นตัวประเมินประสิทธิภาพที่ดีกว่าค่าความแม่นยำและหากสังเกตที่ความสามารถในการจำแนกกรณีที่ข้อมูลตัวอย่างสมดุลเทียบกับกรณีที่ข้อมูลตัวอย่างไม่สมดุลจะเห็นได้ว่า เมื่อข้อมูลตัวอย่างเกิดความไม่สมดุลขึ้นในจำนวนตัวอย่างที่เท่ากัน (เช่น จำนวนตัวอย่างเท่ากับ 12 ระดับความสามารถในการจำแนกของชุดข้อมูลตัวอย่างที่สมดุลเท่ากับ 16.3 แต่ความสามารถในการจำแนกของชุดข้อมูลตัวอย่างที่ไม่สมดุลเท่ากับ 18.4) จะเห็นได้ว่าชุดข้อมูลตัวอย่างที่ไม่สมดุลนั้นค่า AUC มีความสามารถในการจำแนกสูงกว่าค่าความแม่นยำ

ตาราง ที่ 9 ความ สอดคล้อง และ ความ สามารถ ใน การ จำแนก ของ ค่า AUC และ ความแม่นยำ(ไม่สมดุล ขนาดข้อมูล 10 ตัวอย่าง)

คลาสบวก	คลาสลบ	C	D
1	9	1.0	$\infty$
2	8	0.926	22.3
3	7	0.939	15.5
4	6	0.956	14.9
5	5	0.963	15.2

และสุดท้ายเป็นการทดลองในหลายๆ อัตราส่วนของคลาสบวกและคลาสลบ โดยกำหนดให้มีข้อมูลทั้งหมด 10 ตัวอย่าง โดยเริ่มจากสมดุลคือมีทั้งหมดอย่างละ 5 ตัวอย่างจากนั้นเพิ่มและ ลดคลาสใดคลาสหนึ่งไปเรื่อยๆ จนไม่สามารถลดได้ ในกรณีนี้คือเหลือตัวเดียว ผลลัพธ์จากการทดลองแสดงดังตารางที่ 9

จากการทดลองทั้งสองไม่ว่าเป็นข้อมูลทั้งแบบที่สมดุลและ ไม่สมดุลก็ตาม ผลการทดลองยังคงเป็นไปในทิศทางเดียวกัน ทั้งในมุมมองความสอดคล้องที่ยังคงสอดคล้องกันสูง และในมุมมองความสามารถในการจำแนกที่ AUC มีความสามารถในการจำแนกสูง ขึ้นเรื่อยๆ ตามขนาด ของ ข้อมูล และ ยังมี ความสามารถในการจำแนกสูงมากขึ้นเมื่อข้อมูลเกิดความไม่สมดุลของทั้งสองคลาส

## 4 การประยุกต์ใช้

จากการทดลองที่ผ่านมาได้เปรียบเทียบตัวประเมินประสิทธิภาพทั้งสองคือค่า AUC และ ค่าความแม่นยำ โดยค่า AUC มีประสิทธิภาพดีกว่าค่าความ

แม่นยำ แต่อย่างไรก็ตามในการใช้งานจริงทั้ง AUC และความแม่นยำไม่ใช่เป้าหมายสุดท้าย เช่น ธนาคาร หรือ บริษัทประกันภัย อาจจะมีข้อมูลของลูกค้าอยู่มหาศาลโดยสิ่งที่ต้องการสุดท้ายคือการคาดการณ์การทำให้กับ บริษัท

สมมติว่าข้อมูลของลูกค้ามีการเก็บด้วยแอตทริบิวต์จำนวนหนึ่งและลูกค้าแต่ละรายอาจเป็นผู้ซื้อหรือไม่ใช่ผู้ซื้อผลิตภัณฑ์บางอย่างเนื่องจากปัญหานี้เป็นปัญหาการจำแนกแบบไบนารี ลูกค้าจะได้รับการติดต่อจากแคมเปญการส่งเสริมการขายสำหรับลูกค้าแต่ละรายโดย บริษัทต้องคาดการณ์ว่าในสินค้าชนิดๆ หนึ่งนั้นลูกค้าแต่ละรายมีความต้องการสินค้านั้นมากเพียงใด และ ต้องเพิ่มโอกาสการซื้ออย่างน้อยเพียงใด

อย่างไรก็ตามการประยุกต์ใช้ บริษัท อาจต้องการโปรโมตเพียงเล็กน้อยให้กับลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงที่สุดที่คาดการณ์ไว้ และต้องโปรโมตมากขึ้นสำหรับลูกค้าที่มีแนวโน้มจะซื้อสินค้าลดลง ซึ่งทำให้กำไรที่ได้ต่อลูกค้าแต่ละคนนั้นต่างกันไปด้วยซึ่งในความเป็นจริง ก่อให้เกิดผลดีต่อรายได้ของบริษัทเพราะสามารถลดการโปรโมตเกินจำเป็นสำหรับลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงๆ อยู่แล้ว เช่น ลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงสุด 10% แรกนั้น อาจจะเป็นลูกค้าที่มีการซื้อสินค้าเป็นประจำในการโปรโมตสินค้าที่ลูกค้ากลุ่มนี้ซื้อเป็นประจำอยู่แล้วอาจไม่จำเป็น และเพิ่มโอกาสให้ลูกค้าที่มีแนวโน้มจะซื้อสินค้าลดลงมีโอกาสซื้อสินค้ามากขึ้นด้วย

## 5 สรุป

ในบทความนี้ให้คำจำกัดความอย่างเป็นทางการเกี่ยวกับความสอดคล้อง และ ความสามารถในการจำแนก เพื่อใช้ประเมินผลสำหรับขั้นตอนวิธีการเรียนรู้ในปัญหาการจำแนก กำหนดรูปแบบและเกณฑ์ที่ใช้สำหรับการเปรียบเทียบตัวประเมินประสิทธิภาพทั้งสอง และแสดงให้เห็นอย่างชัดเจนว่า AUC นั้นเป็นตัวประเมินประสิทธิภาพที่ดีกว่าค่าความแม่นยำ และได้นำไปเปรียบเทียบกับเหตุการณ์จริงในธุรกิจเพื่อแสดงผลลัพธ์ที่น่าสนใจว่า AUC เกี่ยวข้องโดยตรงกับกำไรสุทธิมากกว่าค่าความแม่นยำในการตลาดทางตรง

การสร้างแบบจำลองที่มีค่า AUC สูงจึงเป็นที่ต้องการมากกว่าการสร้างแบบจำลองที่มีความแม่นยำสูง

## ข้อมูลอ้างอิง

- [1] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270, 1999.
- [2] F. Provost and T. Fawcett. Analysis and visualization of classifier performance: comparison under imprecise class and cost distribution. In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, pages 43–48. AAAI Press, 1997.
- [3] J. Huang and Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. *Classification algorithms*, pages 299 – 310, 2005
- [4] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145–1159, 1997.
- [5] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982.

- [6] I. Kononenko. Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition. In B. Wielinga, editor, *Current Trends in Knowledge Acquisition*. IOS Press, 1990.
- [7] D.J. Hand and R.J. Till. A simple generalisation of the area under the ROC curve for multipleclass classification problems. *Machine Learning*, 45:171–186, 2001.