

## บทที่ 3

### วิธีการดำเนินการวิจัย

งานวิจัยนี้ศึกษาการสร้างแบบจำลองการทำงานของสมองขณะพิมพ์ โดยเครื่องมือที่งานวิจัยนี้เลือกใช้ได้แก่ การถอดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์พี นอกจากนี้ทางผู้วิจัยยังได้หาเทคนิคต่าง ๆ เพื่อช่วยในการวิเคราะห์ข้อมูลเวลาการพิมพ์ที่มีลักษณะพิเศษคือ ข้อมูลที่มีการเปลี่ยนแปลงตามลำดับเวลา โดยเทคนิคที่งานวิจัยนี้เลือกใช้ ได้แก่ เทคนิคแบ่งกลุ่มข้อมูล และเทคนิคหน้าต่างเลื่อน โดยที่เทคนิคการแบ่งกลุ่มข้อมูลที่งานวิจัยนี้เลือกใช้มี 2 เทคนิคคือ การแบ่งกลุ่มแยกตามแถวของคู่อักขระ และค่าเฉลี่ยเคกลุ่ม ซึ่งในส่วนนี้อธิบายถึงรายละเอียดของ ขอบเขตการทดลอง เครื่องมือและซอฟต์แวร์ที่ใช้ในการพัฒนา ขั้นตอนของการทดลอง และวิธีการวัดผลทดลอง แสดงให้เห็นดังต่อไปนี้

#### 3.1 ขอบเขตการทดลอง

##### 3.1.1 แบบเรียนที่ใช้ทดลอง

แบบเรียนที่งานวิจัยนี้เลือกมาทดลองให้ผู้ฝึกพิมพ์ และเก็บข้อมูลเวลาการพิมพ์ของคู่อักขระ ซึ่งจะนำมาใช้วิเคราะห์สร้างแบบจำลอง โดยแบบเรียนที่นำมาให้ผู้ฝึกนั้นเป็นแบบเรียนที่ได้รับการยอมรับ และใช้ในการเรียนการสอนพิมพ์ติดตามมาตรฐานของกระทรวงศึกษาธิการ ซึ่งเป็นแบบเรียนที่ต้องการให้ผู้ฝึกสามารถจดจำตำแหน่งของแป้นพิมพ์ และพัฒนาทักษะให้มีความรวดเร็วและถูกต้อง บทเรียนที่ใช้ทดลองในงานวิจัยนี้ไม่ครอบคลุมบทเรียนที่มีอักขระที่ต้องกดแป้นยกแคร่ (Shift Key) (เพียงภรณ์, 2550)

##### 3.1.2 กลุ่มผู้ทดลอง

ผู้ทดลองเป็นนักศึกษาระดับปริญญาตรีปี 2 ทั้งหมด 5 คน ฝึกพิมพ์ตามแบบฝึกที่กำหนดและจับเวลาการพิมพ์ของคู่อักขระในแต่ละคู่ตามวิธีจัดเก็บค่าเวลาการพิมพ์ (Keystroke Interval Time) โดยการเก็บข้อมูลนั้นจะให้ผู้ฝึกพิมพ์ ครั้งละ 15 นาที หยุดพัก 5 นาที และฝึกพิมพ์ต่อจนครบ 4 ครั้ง

### 3.2 เครื่องมือและซอฟต์แวร์ที่ใช้ในการพัฒนาระบบ

เครื่องมือ และซอฟต์แวร์ที่นำมาใช้ในการพัฒนาแบบจำลองสมองขณะพิมพ์แสดง รายละเอียดดังตารางนี้

ตารางที่ 3.1

รายละเอียดของเครื่องมือ และซอฟต์แวร์ที่ใช้ในการพัฒนาระบบ

ฮาร์ดแวร์ (Hardware)	Intel® Core™ 2 processor T5500 (1.66 GHz) 1024 MB DDR2
ระบบปฏิบัติการ (Operating System)	Window XP Professional
ซอฟต์แวร์ (Software)	WEKA 3.5.7

### 3.3 ขั้นตอนการทดลอง

#### 3.3.1 การออกแบบสร้างแบบจำลอง

งานวิจัยนี้ได้ออกแบบแบบจำลองที่จะใช้ทดลอง เพื่อวัดประสิทธิภาพการทำงานของเครื่องมือ และเทคนิคต่าง ๆ ที่เลือกใช้ โดยได้แบ่งการทดลองออกเป็น 3 ส่วน ได้แก่ การทดลองวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้นด้วยเครื่องมือต่าง ๆ การทดลองวัดประสิทธิภาพของแบบจำลองที่เพิ่มเทคนิคแบ่งกลุ่มข้อมูล และการทดลองวัดประสิทธิภาพของแบบจำลองที่เพิ่มเทคนิคหน้าต่างเลื่อน ซึ่งขั้นตอนในการสร้างแบบจำลองแต่ละแบบมีความแตกต่างกัน ซึ่งจะแสดงให้เห็นดังต่อไปนี้

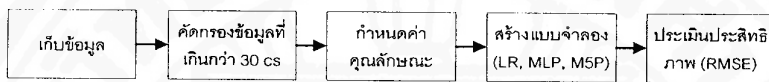
##### 3.3.1.1 การสร้างแบบจำลองแบบที่ 1

แบบจำลองรูปแบบที่ 1 เป็นแบบจำลองเดี่ยวที่อยู่ในตัวแทนชาวนวลาคือ ข้อมูลที่ใช้ในการสร้างแบบจำลองจะไม่มีการจัดการข้อมูลใด ๆ เลย ข้อมูลจะได้มาจากการจัดเก็บ คัดกรอง และกำหนดค่าคุณลักษณะ ต่อมาจึงสร้างแบบจำลองด้วยเครื่องมือทั้ง 3 คือ การถดถอยเชิงเส้น

โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์ที หลังจากนั้นจึงประเมินค่าที่ได้จากการทำนายข้อมูลในกลุ่มทดสอบ และหาค่าเฉลี่ยความผิดพลาดกำลังสองของแบบจำลองที่สร้างด้วยเครื่องมือทั้ง 3 ซึ่งขั้นตอนการทดลองแสดงให้เห็นดังภาพที่ 3.3

ภาพที่ 3.1

ขั้นตอนการทดลองสร้างแบบจำลองแบบที่ 1

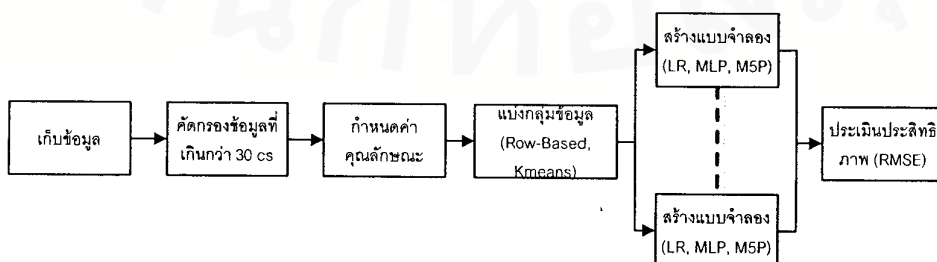


### 3.3.1.2 การสร้างแบบจำลองแบบที่ 2

แบบจำลองรูปแบบที่ 2 นั้นจะแตกต่างกันจากแบบจำลองในรูปแบบที่ 1 คือ เพิ่มเทคนิคการแบ่งกลุ่มข้อมูลก่อนที่จะนำข้อมูลเวลาการพิมพ์ไปวิเคราะห์ด้วยการถดถอยเชิงเส้นโครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์ที ซึ่งเทคนิคการแบ่งกลุ่มที่งานวิจัยนี้เลือกใช้มี 2 เทคนิค ได้แก่ การแบ่งกลุ่มแยกตามแถวของคู่อักขระที่จะแบ่งกลุ่มข้อมูลออกเป็น 7 กลุ่มตามค่าคุณลักษณะแถวของคู่อักขระ และอีกเทคนิคการแบ่งกลุ่มหนึ่งคือ ค่าเฉลี่ยเคกลุ่ม ซึ่งจะแบ่งกลุ่มข้อมูลออกเป็น 10 กลุ่ม ซึ่งในขั้นการวิเคราะห์ข้อมูลจะใช้เครื่องมือตามจำนวนของกลุ่มที่ถูกแบ่งออกตามเทคนิคแบ่งกลุ่มทั้ง 2 วิธี หลังจากนั้นประเมินค่าประสิทธิภาพการทำนายของแบบจำลองด้วยค่าเฉลี่ยความผิดพลาดกำลังสองของแบบจำลองทั้ง 6 แบบที่สร้างขึ้นด้วยการแบ่งกลุ่มข้อมูลแยกตามแถวของคู่อักขระกับเครื่องมือทั้ง 3 และการแบ่งกลุ่มด้วยค่าเฉลี่ยเคกลุ่มกับเครื่องมือทั้ง 3 ซึ่งขั้นตอนการทดลองแสดงให้เห็นดังภาพที่ 3.4

ภาพที่ 3.2

ขั้นตอนการทดลองสร้างแบบจำลองรูปแบบที่ 2

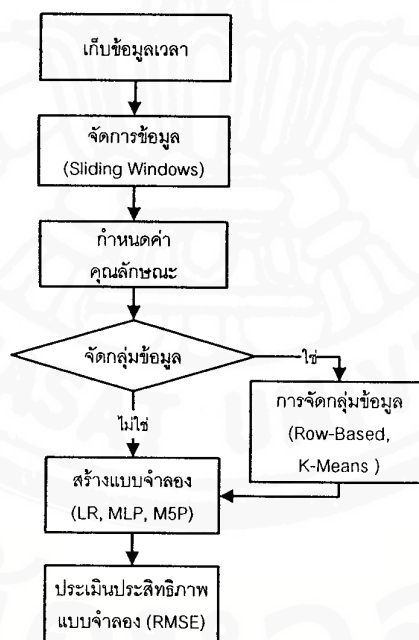


### 3.3.1.3 การสร้างแบบจำลองแบบที่ 3

การทดลองในรูปแบบที่ 3 นี้เป็นการทดลองคล้ายคลึงกับทั้ง 2 แบบที่เสนอมานี้ข้างต้นคือ การนำเอาการทดลองรูปแบบที่ 1 และ 2 มาใช้ทดลองแต่ได้เพิ่มในส่วนของการจัดการข้อมูลด้วยเทคนิคหน้าต่างเลื่อนเข้าไป เพื่อเปรียบเทียบประสิทธิภาพการทำนายของแบบจำลองที่มีรูปแบบเปลี่ยนแปลงไปตามลำดับเวลา (Time series Data) จะมีการทดลอง 9 ครั้งคือ การทดลองด้วยแบบจำลองเดี่ยวจำนวน 3 ครั้งตามจำนวนของเครื่องมือ และการทดลองในรูปแบบที่ 2 ซึ่งมีทั้งหมด 6 ครั้งตามจำนวนของเทคนิคการแบ่งกลุ่มที่เลือกใช้ หลังจากสร้างแบบจำลองเสร็จแล้วประเมินประสิทธิภาพการทำงานของแบบจำลองทั้ง 9 แบบจำลอง ซึ่งขั้นตอนการทดลองแสดงให้เห็นดังภาพที่ 3.3

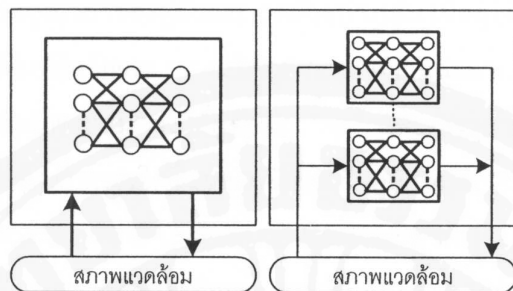
ภาพที่ 3.3

ขั้นตอนการทดลองสร้างแบบจำลองแบบที่ 3



แบบจำลองที่งานวิจัยนี้สร้างนั้นจะมีอยู่ 2 ลักษณะคือ แบบจำลองเดี่ยวที่มีแบบจำลองเดี่ยวในตัวแทนชาญฉลาดที่รับส่งข้อมูลกับสภาพแวดล้อม และแบบจำลองจะมีหลายแบบจำลองอยู่ในตัวแทนชาญฉลาด ซึ่งจะเลือกข้อมูลเข้าวิเคราะห์ในแบบจำลองแต่ละตัวที่มีคุณลักษณะที่แตกต่างกัน

ภาพที่ 3.4  
รูปแบบของแบบจำลองที่ถูกพัฒนา

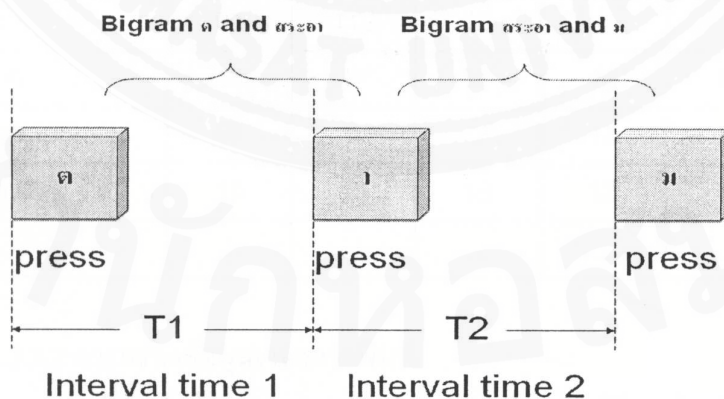


3.3.2 การจัดเตรียมข้อมูล

3.3.2.1 ข้อมูลเวลาการพิมพ์ของผู้ใช้

ค่าเวลาการพิมพ์ (Keystroke Interval Time) เป็นค่าเวลาที่ได้จากการที่ผู้ใช้พิมพ์ตัวอักษรที่เห็นจากแบบเรียน เมื่อผู้ใช้ฝึกพิมพ์กับแบบเรียนที่งานวิจัยนี้กำหนดไว้ ซึ่งค่าดังกล่าวจะเป็นค่าเวลาที่ผู้ใช้แต่ละคนทำได้ตามประสิทธิภาพการพิมพ์ที่ได้เรียนรู้จากแบบเรียน ซึ่งค่าดังกล่าวเป็นค่าพฤติกรรมของผู้ใช้แต่ละคนขณะพิมพ์

ภาพที่ 3.5  
ระยะเวลาการพิมพ์



จากภาพที่ 3.5 แสดงวิธีการจัดเก็บข้อมูลเวลาการพิมพ์ที่ได้จากการฝึกพิมพ์ของผู้ใช้ กับแบบฝึก ซึ่งจะเริ่มจัดเก็บข้อมูลเวลาตั้งแต่ผู้ใช้เริ่มพิมพ์ตัวอักษรแรกไป จนกระทั่งผู้ใช้พิมพ์ตัวอักษรตัวถัดไปจะได้ข้อมูลเวลาของคู่อักษร ยกตัวอย่างเช่น เมื่อผู้ใช้พิมพ์ตัวอักษร “ต” จะเริ่มจับเวลาจนกระทั่งผู้ใช้พิมพ์ตัวอักษร “า” จะได้คู่อักษร “ตา” และค่าเวลาของคู่อักษรดังกล่าวคือ T1 และเมื่อผู้ใช้พิมพ์ตัวอักษร “ม” จะได้คู่อักษร “าม” และค่าเวลาของคู่อักษรดังกล่าวคือ T2 นอกจากนี้ยังได้จัดเก็บลำดับการเกิดขึ้นของคู่อักษรว่าคู่อักษรดังกล่าวเกิดขึ้นเป็นตัวที่เท่าไรของคู่อักษรนั้น ทั้งนี้ค่าระยะเวลาที่จัดเก็บหน่วยเป็นเซนติวินาที (Centisecond: cs) ซึ่งรูปแบบการเก็บค่าเวลา และลำดับการเกิดของคู่อักษรแสดงให้เห็นดังตารางข้างล่างนี้

ตารางที่ 3.2  
ระยะเวลาการพิมพ์ของผู้ใช้

คู่อักษร	เวลาที่พิมพ์ครั้งที่ (cs)					
	ตัวที่ 1	ตัวที่ 2	ตัวที่ 3	ตัวที่ 4	ตัวที่ 5	...
ดต	9	21	20	18	14	...
ดฟ	77	14	14	14	13	...
ฟฟ	18	17	20	17	16	...
ฟว	91	9	12	65	29	...
วว	16	18	18	16	16	...
วก	91	112	...	...	...	...
กก	14	17	17	16	17	...
กา	87	4	9	14	9	...
วา	15	15	15	17	17	...
...	...	...	...	...	...	...

### 3.3.2.2 การกรองข้อมูลเวลาการพิมพ์

ข้อมูลเวลาที่จัดเก็บข้อมูลเวลาการพิมพ์จากแบบเรียนด้วยวิธีการข้างต้นแล้วนั้น ก่อนที่จะนำข้อมูลดังกล่าวมาใช้วิเคราะห์สร้างแบบจำลองการพิมพ์ของผู้ใช้แต่ละคนนั้น ข้อมูลเวลาดังกล่าวจะมีค่าความผิดพลาดที่เกิดขึ้นระหว่างที่ผู้ใช้ฝึกฝนกับแบบเรียนบนอยู่ในกลุ่มข้อมูลเวลา



ที่จัดเก็บ ดังนั้นทางผู้วิจัยจึงคัดกรองข้อมูลเพื่อขจัดข้อมูลที่มีความผิดพลาดออกจากกลุ่มข้อมูล เวลาที่ใช้ทดลอง ซึ่งค่าความผิดพลาดดังกล่าวนั้นจะเป็นข้อมูลเวลาการพิมพ์ที่มีค่าเกินกว่า 30 เซนติวินาที โดยปัจจัยที่ส่งผลให้ค่าเวลาการพิมพ์มีค่ามากกว่า 30 เซนติวินาที นั้นมีสาเหตุอยู่ 3 ประการคือ

- ระยะเวลาที่เกิดจากการที่ผู้ฝึกค้นหาตัวอักษรบนแป้นพิมพ์
- ช่วงที่ผู้ฝึกเสียสมาธิ หรือขาดความสนใจในการพิมพ์ เนื่องจากมีสิ่งรบกวนจากภายนอก เช่น โทรศัพท์ดัง เป็นต้น
- เมื่อผู้ฝึกพิมพ์ตัวอักษรผิด และไม่ได้แก้ไขพิมพ์ตัวอักษรให้ถูกต้องส่งผลให้ข้อมูลที่จัดเก็บมาผิดพลาด

จากการคัดกรองข้อมูลที่มีความผิดพลาดในช่วงต้น การนำข้อมูลดังกล่าวมาใช้เป็นข้อมูลอินพุตในการสร้างแบบจำลองการพิมพ์ด้วยเครื่องมือที่งานวิจัยนี้เลือกใช้ คือ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์พี รวมไปถึงการทดลองที่เพิ่มการแบ่งกลุ่มข้อมูล และเทคนิคหน้าต่างเลื่อน หลังจากคัดกรองข้อมูลแล้วจึงนำข้อมูลดังกล่าวไปกำหนดค่าคุณลักษณะที่ใช้เป็นองค์ประกอบร่วมกับค่าเวลาการพิมพ์ เพื่อวิเคราะห์สร้างแบบจำลองทำนายระยะเวลาการพิมพ์ของผู้ใช้แต่ละคน

### 3.3.2.3 การกำหนดค่าคุณลักษณะของข้อมูล (Attribute)

การจัดรูปแบบของกลุ่มตัวอย่างที่ใช้ข้อมูลให้อยู่ในการทดลองด้วยวิธีการจัดการข้อมูล ช่วงต้น โดยนำข้อมูลดังกล่าวมากำหนดคุณลักษณะของข้อมูลอินพุตก่อนที่จะนำข้อมูลวิเคราะห์สร้างแบบจำลองตามเครื่องมือต่าง ๆ ที่งานวิจัยนี้เลือกใช้ ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์พี ซึ่งในการกำหนดค่าคุณลักษณะที่งานวิจัยนี้เลือกใช้นั้นจะกำหนดตามองค์ประกอบที่มนุษย์ที่ใช้ในการพิมพ์ เช่น นิ้วที่ใช้ในการพิมพ์ ตำแหน่งตัวอักษรบนแป้นพิมพ์ และความสามารถในทักษะการเรียนรู้ของมนุษย์ จากสิ่งที่กล่าวมาในช่วงต้นนี้ผู้วิจัยจึงใช้การกำหนดค่าคุณลักษณะ

การกำหนดค่าคุณลักษณะเป็นการกำหนดในส่วนของข้อมูลส่วนหนึ่งนั้นได้มาจากการศึกษางานวิจัยที่เกี่ยวข้อง และอีกส่วนหนึ่งทางผู้วิจัยได้กำหนดค่าดังกล่าวขึ้นมาเอง และได้ทดลองจากค่าคุณลักษณะที่กำหนดขึ้นมานั้น ด้วยการทดลองตัดค่าคุณลักษณะดังกล่าวออกไปจากข้อมูลอินพุตที่นำไปสร้างแบบจำลอง พบว่าการมีค่าคุณลักษณะดังกล่าวอยู่ทำให้ประสิทธิภาพการทำนายของแบบจำลองดีกว่าแบบจำลองที่ไม่ได้ใส่ค่าคุณลักษณะ

- ตำแหน่งของนิ้วที่ใช้กดเป็นตัวอักขระแรกในคู่อักขระ
- ตำแหน่งของนิ้วที่ใช้กดเป็นตัวอักขระที่สองในคู่อักขระ

ค่าคุณลักษณะทั้งสองนี้บ่งบอกถึงนิ้วที่ใช้ในการกดแป้นพิมพ์ เพื่อพิมพ์ตัวอักขระที่เห็นจากแบบเรียน ตามที่ได้ศึกษาจากแบบจำลองพีดีทีที่กำหนดการกดแป้น และการเคลื่อนที่ไปกดแป้นพิมพ์ของตัวอักขระปลายทาง ซึ่งค่าดังกล่าวจะผ่านการแยกแยะว่าตัวอักขระในคู่อักขระทั้งสองนั้นเป็นตัวอักขระใด

- ลำดับการปรากฏขึ้นของคู่อักขระ

ค่าคุณลักษณะที่บ่งบอก และศึกษาในส่วนของจากสมมติฐานว่า เมื่อจำนวนครั้งในการฝึกพิมพ์คู่อักขระมากขึ้นทักษะ และพัฒนาการในการพิมพ์คู่อักขระ ค่าคุณลักษณะนี้แสดงถึงพัฒนาการของผู้ฝึก ซึ่งได้มาจากการศึกษาในเรื่องของการทำงานของสมองมนุษย์ (อุบลรัตน์, 2535) ที่ทักษะการพิมพ์จะเพิ่มขึ้นจากค่าประสบการณ์ และทักษะนั้นส่งผลให้ประสิทธิภาพการทำงานดีขึ้น

- ความห่างระหว่างตัวอักขระตัวแรกกับตัวอักขระตัวที่สองของคู่อักขระ

ค่าความห่างนั้นจะแสดงถึงระยะทางในการเคลื่อนที่ของนิ้ว ระยะทางในการออกจากตำแหน่งที่นิ้วประจำอยู่ เพื่อพิมพ์ตัวอักขระนั้นระยะทางการเคลื่อนที่นี้เป็นตัวแปรที่ส่งผลต่อความเร็วและความถูกต้องในการพิมพ์ ซึ่งก็ได้มาจากการศึกษาของแบบจำลองพีดีทีอีกส่วนหนึ่งคือ ค่าระยะห่างจากจุดเริ่มต้นไปยังจุดปลายทาง

- ตำแหน่งแถวของตัวอักขระแรกบนแป้นพิมพ์
- ตำแหน่งแถวของตัวอักขระที่สองบนแป้นพิมพ์

ค่าคุณลักษณะทั้งสองนี้บ่งบอกถึงตำแหน่งแถวของตัวอักขระแต่ละตัวในคู่อักขระคือ ตำแหน่งที่ตัวอักขระตัวแรก และตัวที่สองประจำอยู่บนแป้นพิมพ์ช่วยในการแบ่งกลุ่มข้อมูล ซึ่งการแบ่งกลุ่มข้อมูลด้วยเทคนิคแยกตามแถวของคู่อักขระ โดยที่การกำหนดค่าคุณลักษณะดังกล่าว ผู้วิจัยได้ทดลองวัดประสิทธิภาพเมื่อกำหนด และไม่กำหนดคุณลักษณะดังกล่าว ผลที่ได้ปรากฏว่าการกำหนดคุณลักษณะส่งผลให้ประสิทธิภาพการทำงานរបแบบจำลองดีขึ้น

- ค่าคุณลักษณะการกดแป้นยกแคร่ (Shift Key)

ค่าคุณลักษณะนี้แสดงถึงการกดแป้นยกแคร่ของตัวอักขระที่ต้องการการกดแป้นยกแคร่แต่ละตัว เพื่อบ่งบอกถึงการกดแป้นยกแคร่ส่งผลให้ประสิทธิภาพ และค่าเวลาดังกล่าวส่งผลให้เวลาการพิมพ์ลดลง



- ค่าเวลาการพิมพ์ ระยะเวลาที่ได้จากการเก็บข้อมูล

ค่าคุณลักษณะที่บ่งบอกถึงข้อมูลเวลาที่ผู้ใช้ทำได้ และค่าคุณลักษณะเอาต์พุตที่แสดงค่าเวลาการพิมพ์ที่ได้จากการทำนายของแบบจำลอง เพื่อใช้เป็นข้อมูลอินพุตในการเรียนรู้ของเครื่องมือสร้างแบบจำลอง เพื่อให้ได้เครื่องมือที่มีความสามารถทำนายข้อมูลเวลาการพิมพ์ของผู้ใช้แต่ละคนได้

### 3.3.2.4 การแบ่งกลุ่มข้อมูล (Clustering Data)

จากที่ได้ข้อมูลจากที่ได้เก็บข้อมูล คัดกรอง และกำหนดค่าคุณลักษณะด้วยวิธีการข้างต้นมาแล้วนั้น ก่อนที่จะสร้างแบบจำลองรูปแบบที่ 2 ที่มีการเพิ่มเทคนิคแบ่งกลุ่มข้อมูลสร้างแบบจำลองด้วย การถดถอยเชิงเส้น โคจรข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์นั้น การแบ่งกลุ่มข้อมูลในงานวิจัยนี้ใช้ ได้แก่ การแบ่งกลุ่มแยกตามแถวของคู่อักขระ และการแบ่งกลุ่มด้วยค่าเฉลี่ยเคกลุ่ม ซึ่งการแบ่งกลุ่มดังกล่าวอาจช่วยเพิ่มประสิทธิภาพในการวิเคราะห์ข้อมูลที่มีการเปลี่ยนแปลงตามลำดับเวลาขั้นตอน และวิธีการทดลองแสดงให้เห็นดังต่อไปนี้

#### 3.3.2.4.1 การแบ่งกลุ่มข้อมูลแยกตามแถวของคู่อักขระ (Row Based Clustering)

การศึกษานี้เสนอวิธีการแบ่งกลุ่มข้อมูล เพื่อเพิ่มประสิทธิภาพการทำงานของแบบจำลองที่สร้างขึ้นจึงได้เสนอวิธีการแบ่งกลุ่มข้อมูลแยกตามแถวของคู่อักขระ ซึ่งจะมีการแบ่งกลุ่มของข้อมูลที่ผ่านการคัดกรอง และกำหนดค่าคุณลักษณะแล้วแบ่งกลุ่มข้อมูลออกเป็น 7 กลุ่มย่อยตามค่าคุณลักษณะตำแหน่งแถวที่ตัวอักขระแรก และตัวอักขระที่สองบนแป้นพิมพ์

ภาพที่ 3.6

การแบ่งกลุ่มข้อมูลตามตำแหน่งแถวตัวอักขระ



จากภาพที่ 3.6 แสดงให้เห็นถึงตำแหน่งของตัวอักษร และแถวของคู่อักษร แต่ละตัวประจำอยู่ ซึ่งในการแบ่งตามแถวของตัวอักษรแรก กับตัวอักษรที่สองของคู่อักษร ประจำอยู่นั้น ซึ่งจำนวนข้อมูลภายในแต่ละกลุ่มที่แบ่งออกมานั้นจะไม่เท่ากัน ซึ่งก็เป็นไปตาม จำนวนของตัวอักษรในแต่ละแถวไม่เท่ากัน ซึ่งสามารถแบ่งกลุ่มได้ด้วยวิธีการดังต่อไปนี้

- กลุ่มที่ 1 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกอยู่ในแถวที่ 1 กับตำแหน่งของตัวอักษรที่สองอยู่ในแถวที่ 2 และในทางกลับกัน
- กลุ่มที่ 2 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกอยู่ในแถวที่ 1 กับตำแหน่งของตัวอักษรที่สองอยู่ในแถวที่ 3 และในทางกลับกัน
- กลุ่มที่ 3 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกอยู่ในแถวที่ 1 กับตำแหน่งของตัวอักษรที่สองอยู่ในแถวที่ 4 และในทางกลับกัน
- กลุ่มที่ 4 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกอยู่ในแถวที่ 2 กับตำแหน่งของตัวอักษรที่สองอยู่ในแถวที่ 3 และในทางกลับกัน
- กลุ่มที่ 5 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกอยู่ในแถวที่ 2 กับตำแหน่งของตัวอักษรที่สองอยู่ในแถวที่ 4 และในทางกลับกัน
- กลุ่มที่ 6 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกอยู่ในแถวที่ 3 กับตำแหน่งของตัวอักษรที่สองอยู่ในแถวที่ 4 และในทางกลับกัน
- กลุ่มที่ 7 ประกอบไปด้วยกลุ่มของคู่อักษรที่มีตำแหน่งของตัวอักษรแรกและตัวอักษรที่สองอยู่ในแถวเดียวกัน (แถวที่ 1 กับแถวที่ 1, แถวที่ 2 กับแถวที่ 2, แถวที่ 3 กับแถวที่ 3, แถวที่ 4 กับแถวที่ 4)

#### 3.3.2.4.2 การแบ่งกลุ่มข้อมูลด้วยค่าเฉลี่ยเคกลุ่ม (K-Means Clustering)

การแบ่งกลุ่มด้วยค่าเฉลี่ยเคกลุ่มเป็นการแบ่งกลุ่มอีกวิธีที่งานวิจัยนี้เลือกใช้ทดลองที่มีรูปแบบในการแบ่งกลุ่มข้อมูลแตกต่างไปจากการแบ่งกลุ่มตามแถว และเป็นเทคนิคที่นิยมใช้ในการแบ่งกลุ่มข้อมูล เพื่อจัดข้อมูลที่มีค่าคุณลักษณะใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน ซึ่งงานวิจัยนี้ได้แบ่งข้อมูลออกเป็น 10 กลุ่ม และมีขั้นตอนในการแบ่งกลุ่มข้อมูลดังต่อไปนี้

- กำหนดกลุ่มประชากร (k)
- กำหนดค่าของจุดศูนย์กลางของข้อมูลแต่ละกลุ่มโดยสุ่มจากข้อมูลที่มีอยู่
- นำข้อมูลจับเข้าแต่ละกลุ่มที่ได้กำหนดกลุ่มประชากร โดยข้อมูลที่จะจัดเข้ากลุ่มต้องมีค่าใกล้กับจุดศูนย์กลางของกลุ่มข้อมูลมากที่สุด ในแต่ละครั้งที่มีการจับเข้ากลุ่มต้องคำนวณหาค่าจุดศูนย์กลางของกลุ่มข้อมูลนั้นใหม่เสมอ

โดยหาจากค่าเฉลี่ยของข้อมูลในกลุ่ม ซึ่งจะคำนวณจนกระทั่งข้อมูลถูกจัดเข้ากลุ่มจนครบ

- คำนวณสมการวัดค่าความห่างของยูคลิด
- ตรวจสอบผลของสมการวัดค่าความห่างของยูคลิดว่ามีค่าต่ำสุดแทบจะไม่เปลี่ยนแปลง หรือวนรอบตามจำนวนรอบที่กำหนด
- ผลลัพธ์ที่ได้เป็นข้อมูลที่แบ่งเป็นกลุ่มย่อย ๆ จำนวน  $k$  กลุ่ม และมีค่าสมการวัดค่าความห่างของยูคลิดน้อยที่สุด

### 3.3.2.5 การจัดการข้อมูลด้วยเทคนิคหน้าต่างเลื่อน

เทคนิคหน้าต่างเลื่อนเป็นขั้นตอนการจัดการข้อมูล เพื่อช่วยเพิ่มประสิทธิภาพในการทำนายของแบบจำลองต่อข้อมูลเวลาการพิมพ์ที่นำมาทดลอง ซึ่งจะเพิ่มสถานะในการจดจำข้อมูลการพิมพ์ที่เกิดขึ้นมาก่อนหน้า และช่วยปรับค่าข้อมูลที่มีความเปลี่ยนแปลงเป็นอย่างมากให้มีความราบเรียบมากขึ้น ซึ่งขั้นตอนในการจัดการข้อมูลด้วยเทคนิคดังกล่าวแสดงให้เห็นดังสมการต่อไปนี้

$$y = \frac{(x_{i-1} + x + x_{i+1})}{n} \quad (3.1)$$

กำหนดให้

$y$	คือ ข้อมูลเวลาผ่านเทคนิคหน้าต่างเลื่อน
$x$	คือ ข้อมูลเวลาอินพุต
$i$	คือ ตำแหน่งของข้อมูล $i^{th}$ ซึ่ง $i = -1, 0, 1$
$n$	คือ จำนวนของข้อมูล $n = 3$

### 3.3.3 การสร้างแบบจำลอง

ข้อมูลที่ใช้สร้างแบบจำลองได้จากการเก็บเวลาการพิมพ์ คัดกรอง และกำหนดค่าคุณลักษณะ หรือผ่านวิธีการจัดการข้อมูลรูปแบบอื่น เช่น การแบ่งกลุ่มข้อมูล หรือเทคนิคหน้าต่างเลื่อน เพื่อเพิ่มประสิทธิภาพการทำงานของแบบจำลอง และวิเคราะห์ข้อมูลตามสมมุติฐานที่ทางผู้วิจัยได้ตั้งไว้ ซึ่งในขั้นตอนการสร้างแบบจำลองจะมีรูปแบบที่คล้ายคลึงกันแต่จะแตกต่างกันตรงที่เพิ่มส่วนการทำงานบางส่วนเพิ่มเติมเข้าไป ซึ่งค่าข้อมูลที่ป้อนเข้าไปจะอยู่ในรูปแบบเดียวกันคือตามรูปแบบของ WEKA (Witten and Frank, 2005) กำหนดไว้ และเครื่องมือที่ใช้สร้างแบบจำลอง

นั้น WEKA ได้มีไว้ให้เลือกใช้ตามงานที่เหมาะสมซึ่งงานวิจัยนี้เลือกใช้ การถดถอยเชิงเส้น (Linear Regression: LR) โครงข่ายประสาทเทียม (Multilayer Perceptron: MLP) และแบบจำลองต้นไม้เอนไซม์ไฟว์พี (Model Tree: M5P) โดยที่ข้อมูลอินพุตจะถูกกำหนดด้วยค่าคุณลักษณะทั้ง 8 ค่าที่กล่าวถึงมาในข้างต้น ซึ่งรูปแบบของข้อมูลแสดงดังตารางที่ 3.3



ตารางที่ 3.3 ค่าข้อมูลทีป้อนเข้าวิเคราะห์สร้างแบบจำลองด้วย WEKA

คู่อักขระ	ตำแหน่งของนิว คู่อักขระตัวแรก								ตำแหน่งของนิว คู่อักขระตัวแรก								ลำดับคู่อักขระ	ความห่างคู่อักขระ	ตำแหน่งแถวคู่อักขระแรก	กดแป้นยกแคร่	ตำแหน่งแถวคู่อักขระสอง	ค่าเวลาดำเนินการพิมพ์
	มือซ้าย				มือขวา				มือซ้าย				มือขวา									
ดด	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	3	0	3	9
ดด	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	2	0	3	0	3	21
ดฟ	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2	0	3	0	3	14
ดฟ	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	3	0	3	0	3	14
ฟฟ	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	3	0	3	18
ฟว	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	3	0	3	9
วว	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	3	0	3	16
วว	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	2	0	3	0	3	18
กก	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	3	0	3	14

ตารางที่ 3.3 ค่าข้อมูลที่ป้อนเข้าวิเคราะห์สร้างแบบจำลองด้วย WEKA (ต่อ)

คู่อักขระ	ตำแหน่งของนิว คู่อักขระตัวแรก								ตำแหน่งของนิว คู่อักขระตัวแรก								ลำดับคู่อักขระ	ความห่างคู่อักขระ	ตำแหน่งแฉกคู่อักขระแรก	กตแป้นยกแคร์	ตำแหน่งแฉกคู่อักขระสอง	ค่าเวลาการพิมพ์
	มือซ้าย				มือขวา				มือซ้าย				มือขวา									
กา	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	3	0	3	4
กา	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	3	0	3	9
าา	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	3	0	3	15
าา	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2	0	3	0	3	15
...																						



จากรูปแบบข้อมูลข้างต้นจึงแบ่งกลุ่มข้อมูลที่ใช้ทดลองออกเป็น 2 กลุ่มคือ ข้อมูลที่ใช้สำหรับเรียนรู้ (Training Set) และข้อมูลที่ใช้ทดสอบ (Test Set) ตามวิธีการทดสอบประสิทธิภาพแบบไขว้ข้าม 10 กลุ่ม (10 fold Cross validation) ที่จะทดลองทั้งหมด 10 ครั้งด้วยการถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์พี

### 3.3.4 การเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้ของโครงข่ายประสาทเทียมมีขั้นตอนต่าง ๆ ดังแสดงดังต่อไปนี้

- 1) กำหนดค่าน้ำหนักเชื่อมต่อในโครงข่ายเริ่มต้นให้มีค่าน้อย ๆ หรืออาจสุ่มค่าตัวเลขที่มีค่าต่ำ ๆ (-1.0 ถึง +1.0 หรือ -0.5 ถึง +0.5)
- 2) วนซ้ำการทำงาน (ขั้นตอนที่ 3-10) ตรวจสอบเงื่อนไขไม่ถูกต้องจะดำเนินการตามขั้นตอนที่ 3 – 10 ต่อไป
- 3) การฝึกฝน (ขั้นตอนที่ 4 – 10) เพื่อให้โครงข่ายประสาทเทียมได้เรียนรู้ซึ่งประกอบด้วยหลักคือ การเรียนรู้แบบป้อนไปข้างหน้า (ขั้นตอนที่ 3-5) คำนวณค่าความผิดพลาดที่เกิดขึ้นจากวิธีแพร่กระจายย้อนกลับ (ขั้นตอนที่ 6-7) ปรับปรุงค่าน้ำหนัก และค่าการกระตุ้น (ขั้นตอนที่ 8)
- 4) รับสัญญาณอินพุตที่เข้ามาในแต่ละอินพุตโหนด และคำนวณค่าสัญญาณอินพุตของชั้นอินพุต เพื่อส่งต่อไปยังโหนดในชั้นซ่อน
- 5) คำนวณผลรวมค่าน้ำหนัก และค่าอินพุตที่รับเข้ามาจากชั้นอินพุต โดยการคำนวณหาสัญญาณอินพุตของโหนดในชั้นซ่อนนั้นสามารถหาได้จากสมการ (2.1) จากนั้นปรับค่าสัญญาณออกจากโหนดในชั้นซ่อนด้วยฟังก์ชันกระตุ้นแล้วส่งออกเป็นสัญญาณอินพุตของโหนดในชั้นเอาต์พุตดังสมการ (2.2)
- 6) คำนวณผลรวมค่าน้ำหนัก และค่าอินพุตที่รับเข้ามาในชั้นซ่อน ขั้นตอนนี้จะคำนวณคล้ายกับในขั้นตอนที่ 5 โดยคำนวณตามสมการ (2.1) และหาผลรวมค่าสัญญาณอินพุตที่ถูกส่งมาจากชั้นซ่อน ค่าน้ำหนักประจำตัวของอินพุตโหนดในชั้นเอาต์พุตซึ่งเชื่อมต่อมาจากโหนดในชั้นซ่อน จากนั้นปรับค่าสัญญาณออกจากโหนดในชั้นเอาต์พุตด้วยฟังก์ชันกระตุ้นดังสมการ (2.2)
- 7) การหาค่าความคลาดเคลื่อน (Error Signal Term) ในชั้นเอาต์พุตระหว่างค่าที่ได้จริง (Actual Output Values) กับเป้าหมาย (Desired Output Values) เพื่อปรับค่าน้ำหนักโดยสัดส่วนความคลาดเคลื่อนคูณกับค่าอินพุตในชั้นเอาต์พุตดังสมการ

$$\delta_k = \frac{1}{2}(T_k - Z_k)f'(in L_k) \quad (3.2)$$

กำหนดให้

$\delta_k$  คือ ค่าความคลาดเคลื่อนในชั้น k

$T_k$  คือ ค่าเอาต์พุตเป้าหมาย

$Z_k$  คือ ค่าเอาต์พุตที่ได้มาจากชั้น k

$in L_k$  คือ ค่าผลรวมสัญญาณอินพุตของชั้น k

คำนวณค่าความถูกต้องของน้ำหนัก (Weight correction term) เพื่อใช้ในการปรับปรุงน้ำหนักที่เชื่อมต่อไปในชั้นช่อนกับชั้นเอาต์พุตดังสมการ

$$\Delta W_{jk} = \alpha \delta_k Y_j \quad (3.3)$$

กำหนดให้

$\Delta W_{jk}$  คือ ค่าความถูกต้องของน้ำหนักเชื่อมต่อไปในชั้น j กับ k

$\alpha$  คือ ค่าอัตราการเรียนรู้ (Learning Rate)

$\delta_k$  คือ ค่าความคลาดเคลื่อนในชั้น k

$Y_j$  คือ ค่าสัญญาณเอาต์พุตของชั้น j

- 8) หาผลรวมเดลต้าอินพุต (Delta input) ในชั้นช่อนชั้นตอนนี้แตกต่างไปจากในชั้นตอนที่ 7 เนื่องจากชั้นช่อนจะไม่มีค่าเอาต์พุตเป้าหมาย (Output Target) จึงสมการดังต่อไปนี้

$$\delta_j = \sum_{k=1}^N \delta_k W_{jk} f'(in L_j) \quad (3.4)$$

กำหนดให้

$\delta_j$  คือ ค่าความคลาดเคลื่อนในชั้น j

$W_{jk}$  คือ ค่าความถูกต้องของน้ำหนักเชื่อมต่อไปในชั้น j กับ k

$\delta_k$  คือ ค่าความคลาดเคลื่อนในชั้น k

$in L_j$  คือ ค่าผลรวมสัญญาณอินพุตของชั้น j

คำนวณค่าความถูกต้องของน้ำหนักเพื่อใช้ในการปรับปรุงค่าน้ำหนักที่เชื่อมต่อไปในชั้นอินพุตกับชั้นซ่อนดังสมการ (3.3)

9) ปรับปรุงค่าน้ำหนักแต่ละโหนดของชั้นเอาต์พุตด้วยสมการ

$$W_{jk}^t = W_{jk}^{t-1} + \Delta W_{jk} \quad (3.5)$$

กำหนดให้

$W^t$  คือ ค่าน้ำหนักผ่านการปรับค่าในชั้น  $j$  กับ  $k$

$W^{t-1}$  คือ ค่าน้ำหนักก่อนปรับค่าในชั้น  $j$  กับ  $k$

$\Delta W_{jk}$  คือ ค่าความถูกต้องของน้ำหนักเชื่อมต่อไปในชั้น  $j$  กับ  $k$

จากนั้นปรับค่าน้ำหนักใหม่แต่ละตัวในชั้นซ่อนด้วยสมการ (3.5)

10) ทดสอบการทำงานของโครงข่ายประสาทเทียมหลังจากที่ผ่านกระบวนการฝึกฝนเสร็จสิ้นแล้ว หลังจากนั้นจึงป้อนข้อมูลทดสอบเข้าสู่โครงข่ายประสาทเทียมแบบหลายชั้น และคำนวณหาค่าเอาต์พุตจากค่าน้ำหนักที่มีการคำนวณไว้ในขั้นตอนการเรียนรู้

เงื่อนไขการหยุดฝึกกำหนดเงื่อนไขในการหยุดฝึกนั้นสามารถทำได้ 2 กรณีคือ เมื่อผลรวมของค่าผิดพลาดเฉลี่ยระหว่างเอาต์พุตกับเป้าหมายทั้งหมดมีค่าลดลงน้อยกว่าค่าที่ได้จากเอาต์พุตของโครงข่ายประสาทเทียมเทียบกับค่าเป้าหมายที่ใช้สำหรับบ่งชี้ผลของการฝึกฝนเพียงพอกับการใช้งานหรือไม่ อีกกรณีหนึ่งคือ การกำหนดจำนวนรอบของการฝึกฝนให้ดำเนินไปจนถึงค่าที่รอบการเรียนรู้ที่ตั้งไว้ ซึ่งจำนวนที่เหมาะสมได้จากการทดลอง

### 3.4 วิธีวัดผลการทดลอง

กระบวนการตรวจสอบค่าความผิดพลาดในการทำนาย หรือการคำนวณของแบบจำลองที่สร้างขึ้นด้วยวิธีการต่าง ๆ ที่กล่าวถึงในข้างต้น เพื่อวัดประสิทธิภาพในการทำงานของแบบจำลองค่าความผิดพลาดดังกล่าวได้จากการทำนายระยะเวลาการพิมพ์ของแบบจำลองกับค่าระยะเวลาการพิมพ์ที่ผู้ใช้ทำได้นำมาคำนวณตามสมการดังต่อไปนี้

$$RMSE = \sqrt{\frac{\sum_{patterns} (Desired - Actual)^2}{\# patterns}} \quad (3.6)$$

กำหนดให้

**RMSE** (Root Mean Squared Error) คือ ค่าเฉลี่ยของความผิดพลาดที่ได้จากการทำนายของโครงข่ายประสาทเทียม

**Desired** คือ ค่าที่ได้จากโครงข่ายประสาทเทียม

**Actual** คือ ค่าข้อมูลเป้าหมาย

**Pattern** คือ จำนวนของข้อมูลทั้งหมด

แบบจำลองจะทดสอบหาค่าเฉลี่ยความผิดพลาดกำลังสองที่ประเมินประสิทธิภาพด้วยวิธีการทดสอบแบบไขว้ข้ามสิบกลุ่ม ซึ่งการหาค่าความผิดพลาดหลังจากแบ่งเป็น 10 กลุ่ม และนำข้อมูลที่ได้ไปสร้างแบบจำลองจากกลุ่มข้อมูลเรียนรู้ และทดสอบด้วยกลุ่มข้อมูลทดสอบ หลังจากทำนายด้วยแบบจำลองจึงวัดประสิทธิภาพของแบบจำลองด้วยค่าเฉลี่ยความผิดพลาดกำลังสองของทั้ง 10 กลุ่ม เพื่อเปรียบเทียบประสิทธิภาพการทำงานของแบบจำลองที่สร้างขึ้นด้วยการถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์พี นอกจากนี้ยังรวมถึงการนำเอาเทคนิคแบ่งกลุ่มข้อมูลไม่ว่าจะเป็นการแบ่งกลุ่มข้อมูลแยกตามแถวของคู่อักขระ และค่าเฉลี่ยเคกลุ่ม อีกทั้งยังเพิ่มเทคนิคหน้าตาต่างเดือนที่จะจัดการปรับค่าข้อมูลอินพุต เพื่อหาระยะเวลาในการพิมพ์ตัวอักขระ และหาค่าอัตราการเรียนรู้ของผู้ใช้ ซึ่งนำค่าดังกล่าวไปจัดทำกราฟการเรียนรู้การพิมพ์ของคู่อักขระที่บ่งบอกถึงพัฒนาการในการพิมพ์ของผู้ใช้แต่ละคน

### 3.5 สมมติฐานของการวิจัย

1. การทดลองสร้างแบบจำลอง เพื่อเปรียบเทียบประสิทธิภาพของเครื่องมือสร้างแบบจำลองการทำงานของสมองขณะพิมพ์ด้วย การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไพร์พี
2. การทดลองสร้างแบบจำลอง เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองที่ผ่านการแบ่งกลุ่มข้อมูลกับข้อมูลที่ไม่ผ่านการแบ่งกลุ่ม
3. การทดลองสร้างแบบจำลอง เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองที่ผ่านการแบ่งกลุ่มข้อมูลด้วยวิธีแยกตามแถวของคู่อักขระ กับค่าเฉลี่ยเคกลุ่ม

4. การทดลองสร้างแบบจำลอง เพื่อวัดประสิทธิภาพของแบบจำลองด้วยชุดข้อมูลที่ผ่านการจัดการข้อมูลเปลี่ยนแปลงตามลำดับเวลาด้วยเทคนิคหน้าต่างเลื่อน

