

แอปพลิเคชันสำหรับค้นหาสุนัขสูญหาย

APPLICATION FOR FINDING MISSING DOGS



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

แอปพลิเคชันสำหรับค้นหาสุนัขสูญหาย

APPLICATION FOR FINDING MISSING DOGS



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญาโทปีการศึกษา 2559

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง แอปพลิเคชันสำหรับค้นหาสุนัขสูญหาย

APPLICATION FOR FINDING MISSING DOGS

ผู้จัดทำ

1. นายภคิน บุญชูพิริชัย รหัสนักศึกษา 56010901

2. นางสาวศิริพร กาญจนไพจิตร รหัสนักศึกษา 56011216



ศุภณัฐ ศรีปลั่ง อาจารย์ที่ปรึกษา
(อาจารย์ชุตินเมษฐ์ ศรีนิลทา)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แอปพลิเคชันสำหรับค้นหาสุนัขสูญหาย

นายภคิน	บุญชูพิริชัย	56010901
นางสาวศิริพร	กาญจนไพจิตร	56011216
ผศ.ดร.ชุตินเมษฏ์	ศรีนิลทา	อาจารย์ที่ปรึกษา
ปีการศึกษา 2559		

บทคัดย่อ

สุนัขเป็นสัตว์เลี้ยงที่มีความใกล้ชิดและผูกพันกับมนุษย์เหมือนเป็นสมาชิกในครอบครัว การสูญหายของสุนัขทำให้ทั้งตัวสุนัขเองและเจ้าของต้องพบกับความทุกข์และความเจ็บปวด หากสุนัขไม่สามารถหาทางเพื่อกลับไปสู่เจ้าของได้ สุนัขเหล่านั้นจะต้องอาศัยตามท้องถนนและอาจเป็นอันตรายต่อผู้อื่น กลายเป็นปัญหาของชุมชนและสังคม

ปัญหานี้จึงถูกคิดค้นขึ้นเพื่อแก้ปัญหาดังกล่าว กล่าวคือเพื่อให้สุนัขและเจ้าของมีโอกาสได้กลับมาพบกันอีกครั้ง โดยการสร้างแอปพลิเคชันค้นหาสุนัขสูญหายซึ่งทุกคนสามารถให้ความร่วมมือกันช่วยค้นหาสุนัขโดยใช้แอปพลิเคชันนี้ได้ เมื่อสุนัขสูญหายหรือถูกพบ ผู้ใช้สามารถให้ข้อมูลต่าง ๆ กับระบบ เช่น สายพันธุ์ อายุ เพศ จุดสังเกต สี ตำแหน่ง เวลาและรูปภาพ เป็นต้น ด้วยการประยุกต์ใช้วิธีการประมวลผลภาพ การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูลระบบจึงสามารถแนะนำการจับคู่ที่เป็นไปได้ระหว่างสุนัขสูญหายและสุนัขที่ถูกพบโดยใช้ความคล้ายกันของใบหน้าสุนัข ข้อมูลอื่นเช่น สายพันธุ์ อายุ หรือเพศ จะถูกใช้เพื่อคัดกรองจำนวนผลลัพธ์ที่เป็นไปได้ให้ลดลง ซึ่งแอปพลิเคชันจะสามารถลดจำนวนของสุนัขไว้บ้านลงโดยการลดเวลาในการค้นหาและเพิ่มโอกาสในการพบเจอมากขึ้น นอกเหนือไปจากนี้ แอปพลิเคชันยังสามารถให้โอกาสแก่ผู้ใช้ที่ต้องการรับเลี้ยงสุนัขเลือกลักษณะของสุนัขตามที่ต้องการได้อีกด้วย

กิตติกรรมประกาศ

ปริญญาโทฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี ด้วยคำแนะนำและคำปรึกษาจาก ผศ.ดร. ชุติเมษณ์ ศรีนิลทา อาจารย์ที่ปรึกษา ที่คอยให้ความช่วยเหลือ ให้ความรู้ แนะนำแนวทางเมื่อเกิดปัญหาและชี้แนะจุดบกพร่องของปริญญาโทให้สมบูรณ์ยิ่งขึ้น ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากท่านอาจารย์และขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณคณาจารย์ในภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ให้ความรู้ต่าง ๆ ที่เป็นส่วนช่วยในการทำปริญญาโทฉบับนี้ตั้งแต่อดีตมาจนถึงปัจจุบัน

ขอขอบคุณ ห้องวิจัย Software and Application Interest Group ของภาควิชาวิศวกรรมคอมพิวเตอร์ ที่อำนวยความสะดวกในด้านสถานที่และการยืมอุปกรณ์ เครื่องมือ ตลอดจนเพื่อน ๆ พี่ ๆ ทุกคนที่ได้มีส่วนช่วยให้คำปรึกษาและให้กำลังใจเสมอมา

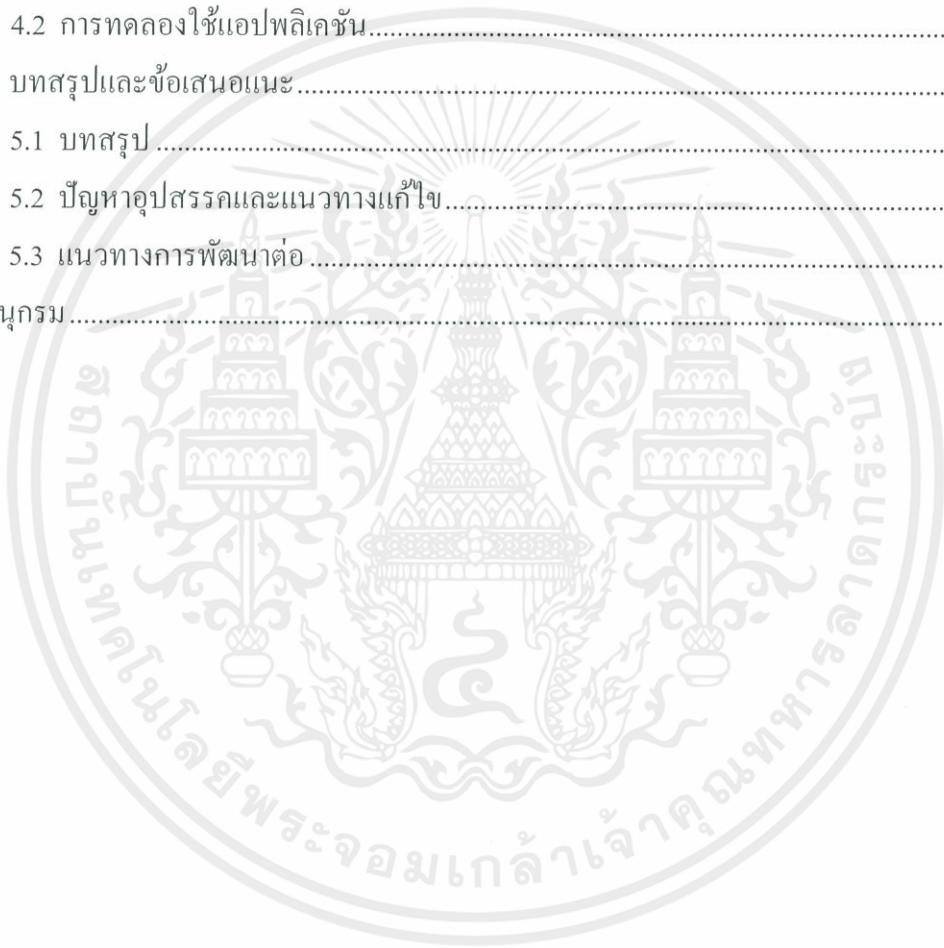
สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้า ผู้ที่คอยให้กำลังใจและจัดหางบประมาณในการจัดทำปริญญาโทฉบับนี้

ภคิน บุญชูพิริชย์
ศิริพร กาญจนไพจิตร

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญรูป.....	VI
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและที่มาของโครงการ.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 ขอบเขตของโครงการ.....	2
1.4 วิธีการดำเนินการ.....	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.6 ข้อยกเว้นของโปรแกรม.....	3
1.7 ส่วนประกอบปริยญาณินพนธ์.....	4
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1 การปรับปรุงคุณภาพของรูปภาพ (Image Enhancement).....	5
2.2 การเรียนรู้ของเครื่อง (Machine Learning).....	6
2.3 การจำแนกข้อมูล (Data Classification).....	7
2.4 การแบ่งกลุ่มข้อมูล (Clustering).....	8
2.5 การคัดเลือกคุณสมบัติของข้อมูล (Feature Selection).....	8
2.6 การค้นหาเพื่อนบ้านใกล้สุด (Nearest Neighbor Search).....	9
2.7 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis).....	9
2.8 การแบ่งกลุ่มด้วยวิธี Affinity Propagation.....	10
2.9 โครงข่ายประสาทเทียม OverFeat.....	11
2.10 การหาระยะทางแบบยูคลิด (Euclidean Distance).....	11
2.11 การแบ่งกลุ่มข้อมูลแบบเคมีน (K-Means Clustering).....	12
2.12 การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis).....	13
2.13 เทคโนโลยีที่ใช้.....	14
2.14 เครื่องมือที่ใช้.....	15

บทที่ 3 การออกแบบและพัฒนา.....	14
3.1 ความต้องการของระบบและฟังก์ชันการทำงาน.....	14
3.2 โครงสร้างของระบบ.....	19
3.3 การออกแบบอัลกอริทึมสำหรับคัดแยกและค้นหาใบหน้าสุนัข.....	20
3.4 การทำงานของอัลกอริทึมค้นหาใบหน้าสุนัข.....	24
3.5 ER diagram ของระบบ.....	25
บทที่ 4 การทดลองและผลการทดลอง.....	27
4.1 การทดลองอัลกอริทึมของระบบ.....	27
4.2 การทดลองใช้แอปพลิเคชัน.....	38
บทที่ 5 บทสรุปและข้อเสนอแนะ.....	44
5.1 บทสรุป.....	44
5.2 ปัญหาอุปสรรคและแนวทางแก้ไข.....	45
5.3 แนวทางการพัฒนาต่อ.....	45
บรรณานุกรม.....	46



สารบัญรูป

รูป	หน้า
รูป 2.1 ขั้นตอนการกรองภาพด้วยการกรองแบบคอนโวลูชัน	6
รูป 2.2 กระบวนการของอัลกอริทึม Nearest Neighbor Search	9
รูป 2.3 การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย.....	10
รูป 2.4 สัญลักษณ์ของ MySQL	14
รูป 2.5 สัญลักษณ์ของ Django.....	14
รูป 2.6 สัญลักษณ์ของภาษา Python.....	15
รูป 2.7 สัญลักษณ์ของภาษา Java.....	15
รูป 2.8 สัญลักษณ์ของ Scikit-learning.....	16
รูป 2.9 สัญลักษณ์ของ Android Studio.....	17
รูป 2.10 สัญลักษณ์ของ JetBrains PyCharm.....	17
รูป 3.1 แผนผังโครงสร้างของระบบ.....	19
รูป 3.2 การจัดเตรียมรูปใบหน้าสุนัข	21
รูป 3.3 ขั้นตอนการสร้างโมเดล	23
รูป 3.4 แผนผังการทำงานของอัลกอริทึมสำหรับค้นหารูปใบหน้าสุนัขเหมือน	24
รูป 3.5 แสดง ER diagram ของระบบ.....	25
รูป 4.1 กระบวนการสร้างโมเดลด้วยอัลกอริทึม Linear Regression Analysis.....	28
รูป 4.2 แสดงการคัดกรองคุณสมบัติและการลดการกระจายค่าของคุณสมบัติในปริภูมิ.....	28
รูป 4.3 กราฟแสดงผลการทดลองการจัดกลุ่มด้วยอัลกอริทึม K-Means.....	31
รูป 4.4 แสดงผลการทดลองเปรียบเทียบ.....	32
รูป 4.5 ผลลัพธ์เมื่อผ่าน Clustering.....	33
รูป 4.6 ผลลัพธ์สุดท้ายของการทดลอง	37
รูป 4.7 สัดส่วนของผู้ทดลองที่ให้คะแนนความพึงพอใจ 31 คน	38
รูป 4.8 ส่วนแสดงผลการเข้าสู่ระบบ	39
รูป 4.9 ส่วนแสดงรายการสุนัขของผู้ใช้.....	40
รูป 4.10 ส่วนแสดงประกาศสุนัขสูญหาย	41
รูป 4.11 ส่วนแสดงประกาศพบเจอสุนัข.....	42

สารบัญญรูป (ต่อ)

รูป 4.12 ส่วนแสดงการค้นหาสุนัขใบหน้าคล้าย.....43



เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของโครงการ

ในปัจจุบันสังคมไทยมีจำนวนสุนัขไร้บ้านเพิ่มมากขึ้น มีสาเหตุหลักมาจากสุนัขสูญหายและถูกทอดทิ้ง ส่งผลให้การขยายพันธุ์ของสุนัขไม่มีการควบคุม จากการสำรวจจำนวนประชากรสุนัขเมื่อปี 2557 ของสำนักงานปศุสัตว์ในประเทศไทย พบว่า มีสุนัขในประเทศไทยมากกว่า 8.5 ล้านตัว ในจำนวนนี้เป็นสุนัขไร้บ้านถึง 700,000 ตัว แบ่งเป็นเพศผู้ 370,000 ตัว เพศเมีย 340,000 คาดการณ์ว่า สุนัขเพศเมีย 1 ตัว มีโอกาสออกลูกได้มากถึง 10 ตัวต่อปี ดังนั้นหากไม่มีการแก้ไขปัญหานี้ สุนัขไร้บ้านจะมีแนวโน้มเพิ่มขึ้นสูงสุดถึง 3,400,000 ตัวต่อปี

จากข้อมูลทางสถิติของ The Humane Society of the United States ซึ่งเป็นองค์กรที่ทำหน้าที่ปกป้องและคุ้มครองสัตว์เลี้ยง มีจำนวนสัตว์เลี้ยงที่สูญหายในระยะเวลา 1 ปี ประมาณ 3.5 ล้านตัว โดย 80 % ของสัตว์เลี้ยงที่สูญหายนี้เป็นสุนัข ซึ่งเป็นสาเหตุที่ทำให้สุนัขเหล่านั้นกลายเป็นสุนัขไร้บ้าน ซึ่งการสูญหายของสุนัขสามารถเกิดขึ้นได้บ่อยครั้ง สาเหตุจากอุปนิสัยชอบอิสระของสุนัข สุนัขเข้าคู่วงของการผสมพันธุ์ เทศกาลต่าง ๆ ที่มีการจุดพลุทำให้สุนัขตื่นกลัวและวิ่งเตลิด หรือลี้ภัยหนีสุนัขไว้เมื่อพาสุนัขออกไปข้างนอกโดยเฉพาะสุนัขที่เป็นพันธุ์ขนาดเล็ก อีกทั้งสุนัขยังถือว่าเป็นสัตว์เลี้ยงที่มีความใกล้ชิดและผูกพันกับมนุษย์จนเหมือนกับเป็นสมาชิกในครอบครัว ดังนั้นเมื่อสุนัขสูญหายจึงทำให้เจ้าของและสุนัขเกิดความทุกข์ซึ่งอาจส่งผลกระทบต่อจิตใจและสุขภาพ

ในส่วนของสุนัขที่ถูกทอดทิ้ง มีสาเหตุมาจากการที่เจ้าของไม่สามารถเลี้ยงได้ เช่น ลูกสุนัขเกิดใหม่จากการที่ไม่ได้คุมกำเนิด สุนัขเป็นโรคซึ่งเป็นที่ยังเกียจหรือไม่สามารถรักษาให้หายขาดได้ หรือภาวะหนี้สินซึ่งไม่อาจแบกรับค่าใช้จ่ายของสุนัขไหว สุนัขเหล่านี้จึงเป็นสาเหตุหนึ่งที่ทำให้มีจำนวนสุนัขไร้บ้านเพิ่มมากขึ้น โดยสุนัขไร้บ้านที่ไม่มีคนรับเลี้ยงเหล่านี้อาจเป็นแหล่งเพาะเชื้อโรคต่าง ๆ ที่จะติดต่อไปสู่มนุษย์ เช่น โรคพิษสุนัขบ้า ปรสิติ แบคทีเรียหรือโรคผิวหนัง เป็นต้น

ในปัจจุบันช่องทางสำหรับการตามหาสุนัขสูญหายและรับอุปการะสุนัขไร้บ้านยังไม่มีประสิทธิภาพเท่าที่ควร จึงทำให้เจ้าของสุนัขใช้เวลาในการค้นหาและใช้กำลังเงินเป็นจำนวนมากในการค้นหาด้วยตนเอง เช่น การสอบถามจากบุคคลในบริเวณที่สุนัขสูญหาย การทำใบติดประกาศในที่สาธารณะหรือการประกาศหาทางอินเตอร์เน็ตและวิทยุต่าง ๆ เป็นต้น ในขณะที่ผู้พบเจอสุนัขก็ไม่ทราบว่าจะจะไปแจ้งที่ใดหรือหน่วยงานไหน ทำให้โอกาสที่เจ้าของจะได้สุนัขกลับคืนน้อยมากและผู้ที่ต้องการรับอุปการะสุนัขมีช่องทางในการค้นหาสุนัขที่มีลักษณะตามที่ต้องการไม่มาก

จากปัญหาดังกล่าว ทางผู้พัฒนาจึงมีแนวคิดที่จะนำเทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning) มาประยุกต์ใช้กับแอปพลิเคชันเพื่อเป็นทางเลือกให้แก่ผู้ใช้ในการค้นหาสุนัขสูญหาย ซึ่งช่วยลดระยะเวลาในการค้นหาและเพิ่มโอกาสในการพบสุนัข อีกทั้งยังเปิดโอกาสให้ผู้ต้องการรับเลี้ยงสุนัขสามารถเลือกสุนัขที่มีลักษณะตรงตามความต้องการ เป็นการลดจำนวนประชากรของสุนัขไร้บ้านได้อีกทางหนึ่ง

1.2 วัตถุประสงค์ของการศึกษา

- 1) พัฒนาแอปพลิเคชันสำหรับช่วยเหลือให้เจ้าของสุนัขมีโอกาสพบสุนัขที่หายไปสูงขึ้น และรวดเร็วยิ่งขึ้น
- 2) พัฒนาแอปพลิเคชันที่ช่วยเพิ่มความสะดวกในการแจ้งข้อมูลการพบเห็นของสุนัขที่หายไป
- 3) พัฒนาแอปพลิเคชันเพื่อเพิ่มโอกาสให้สุนัขไร้บ้านได้รับการอุปการะ

1.3 ขอบเขตของโครงการ

- 1) แอปพลิเคชันสามารถจัดเก็บประวัติและพิกัดปัจจุบันของผู้ใช้และสุนัขได้
- 2) แอปพลิเคชันสามารถเสนอภาพของสุนัขในระบบที่มีความคล้ายคลึงกับภาพของสุนัขสูญหายหรือสุนัขไร้บ้าน
- 3) แอปพลิเคชันสามารถค้นหาสุนัขสูญหายหรือสุนัขไร้บ้านจากพิกัดที่มีผู้พบเห็นสุนัขล่าสุดรวมไปถึงลักษณะที่โดดเด่นของสุนัข
- 4) แอปพลิเคชันสามารถสร้างประกาศหาและประกาศพบสุนัขสูญหายและสุนัขไร้บ้านได้
- 5) แอปพลิเคชันสามารถแจ้งเตือนอัตโนมัติเมื่อผู้พบสุนัขบันทึกภาพสุนัขที่มีลักษณะใกล้เคียงกับสุนัขสูญหายลงในระบบ

1.4 วิธีการดำเนินการ

- 1) กำหนดวัตถุประสงค์และขอบเขตของการทำงาน
- 2) ศึกษาเทคนิคต่าง ๆ ที่สามารถนำมาประยุกต์ใช้กับการตรวจจับใบหน้าของสุนัข (Dog Face Detection)
- 3) ค้นหาเทคนิคที่เหมาะสมในการรู้จำใบหน้าของสุนัข (Dog Face Recognition)
- 4) วิเคราะห์และออกแบบระบบจากการศึกษาหลักการ ทฤษฎี และเอกสารที่เกี่ยวข้อง
- 5) ค้นหาและจัดเตรียมกลุ่มรูปภาพของสุนัข (Dataset) สำหรับใช้ในการทดลอง
- 6) ศึกษาและทดลองเขียน โปรแกรม โดยใช้ภาษา python และใช้ library ที่เหมาะสม
- 7) ศึกษาการเขียนแอปพลิเคชันบนระบบปฏิบัติการแอนดรอยด์ ออกแบบส่วนติดต่อผู้ใช้งาน (User Interface)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้รับความรู้จากการเขียน โปรแกรมในด้านการเรียนรู้ของเครื่อง
- 2) ได้รับความรู้ความเข้าใจในเรื่องของการตรวจจับและรู้จำใบหน้าของสุนัข (Dog Face Detection and Recognition)
- 3) สามารถนำกระบวนการจดจำใบหน้าสุนัข มาประยุกต์ใช้ให้เกิดประโยชน์
- 4) โปรแกรมที่พัฒนาขึ้นสามารถนำไปต่อยอดและใช้ในการค้นหาและรับเลี้ยงสุนัขได้จริง

1.6 ข้อจำกัดของโปรแกรม

- 1) ภาพใบหน้าที่จะนำมาวิเคราะห์ ต้องเป็นภาพหน้าตรงของสุนัขที่เห็นองค์ประกอบบนใบหน้าครบถ้วน
- 2) ไฟล์รูปภาพที่นำเข้าสู่โปรแกรมนั้น ต้องเป็นไฟล์รูปภาพนามสกุล .jpg เท่านั้น
- 3) สภาพแวดล้อมรอบ ๆ จะต้องไม่กลมกลืนไปกับใบหน้าของสุนัข
- 4) รูปภาพที่ป้อนเข้ามาในระบบควรเป็นรูปใบหน้าสุนัขหน้าหนึ่ง ไม่อ้าปาก
- 5) ควรกำหนดสภาพแสงให้มีความเหมาะสม ไม่อยู่ในที่มืด หรือสว่างจนเกินไป

1.7 ส่วนประกอบปริญญาโท

ปริญญาโทฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บท ได้แก่ บทนำ ทฤษฎีที่เกี่ยวข้อง การออกแบบและพัฒนา การทดลองและผลการทดลอง บทสรุปและข้อเสนอแนะ โดยมีรายละเอียดดังนี้

บทที่ 1 บทนำ กล่าวถึงความสำคัญและที่มาของโครงการ วัตถุประสงค์ของโครงการ ขอบเขตของโครงการ วิธีการดำเนินการ ประโยชน์ที่คาดว่าจะได้รับ ข้อจำกัดของ โปรแกรมและ ส่วนประกอบของปริญญาโท

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง กล่าวถึงทฤษฎีที่ได้ศึกษามาใช้ในการออกแบบและพัฒนาระบบ ได้แก่ การปรับปรุงคุณภาพของรูปภาพ (Image Enhancement) การเรียนรู้ของเครื่อง (Machine Learning) การจำแนกข้อมูล (Data Classification) การแบ่งกลุ่มข้อมูล (Clustering) การคัดเลือกคุณสมบัติของข้อมูล (Feature Selection) การค้นหาเพื่อนบ้านใกล้สุด (Nearest Neighbor Search) การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) การแบ่งกลุ่มด้วยวิธี Affinity Propagation โครงข่ายประสาทเทียม Overfeat การหาระยะทางแบบยูคลิด (Euclidean Distance) การแบ่งกลุ่มข้อมูลแบบเคมีน (K-Means clustering) การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) เทคโนโลยีที่ใช้และเครื่องมือที่ใช้

บทที่ 3 การออกแบบและพัฒนา กล่าวถึงภาพรวมและองค์ประกอบของระบบ ได้แก่ ความต้องการของระบบและฟังก์ชันการทำงาน โครงสร้างของระบบ การออกแบบอัลกอริทึมสำหรับคัดแยกและค้นหาใบหน้าสุนัข การทำงานของอัลกอริทึมค้นหาใบหน้าสุนัข

บทที่ 4 การทดลองและผลการทดลอง แบ่งออกเป็น การทดลองอัลกอริทึมของระบบและการทดลองใช้แอปพลิเคชัน

บทที่ 5 บทสรุปและข้อเสนอแนะ กล่าวถึงบทสรุป ปัญหาอุปสรรคและแนวทางแก้ไข และ แนวทางการพัฒนาต่อ

บทที่ 2

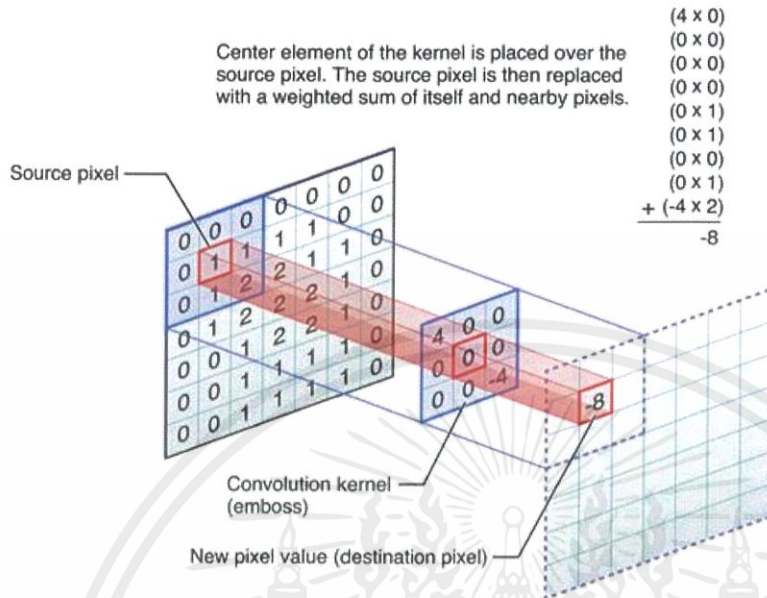
ทฤษฎีที่เกี่ยวข้อง

2.1 การปรับปรุงคุณภาพของรูปภาพ (Image Enhancement)

เป็นกระบวนการในการแปลงข้อมูลภาพตัวเลข เพื่อที่จะสร้างภาพที่เน้นรายละเอียดหรือปรับโทนแสงของภาพที่ต้องการ โดยอาจจะต้องพิจารณาใช้เทคนิคปรับปรุงภาพหลายเทคนิคและนำมาประมวลผลเพื่อให้ได้ภาพที่ต้องการ ซึ่งแต่ละเทคนิคจะมีการปรับปรุงภาพในแง่มุมที่แตกต่างกัน เทคนิคที่ใช้การปรับปรุงภาพวิธีหนึ่งที่เป็นที่นิยมคือ การปรับปรุงภาพเชิงพื้นที่ (Spatial Enhancement) ซึ่งเป็นการเน้นภาพที่ใช้ค่าจากจุดภาพที่อยู่รอบ ๆ จุดภาพใด ๆ มาคำนวณ โดยการปรับภาพเชิงพื้นที่จะเกี่ยวข้องกับความสัมพันธ์เชิงพื้นที่ หมายถึง ความแตกต่างระหว่างค่าสูงสุดและต่ำสุดของกลุ่มจุดบนภาพที่อยู่ติดกันหรือใกล้เคียงกัน เทคนิคที่นิยมใช้งานกัน ได้แก่ การเน้นขอบของวัตถุในภาพ (Edge Enhancement) และการกรองภาพแบบคอนโวลูชัน (Convolution Filtering)

- 1) การเน้นขอบภาพ (Edge Enhancement) เป็นการเน้นขอบโดยทำให้จุดภาพตรงรอยต่อระหว่างพื้นที่ที่มีกลุ่มจุดภาพมีค่าเพิ่มสูงขึ้นหรือลดลง ขึ้นกับความสว่างและความมืดของขอบ หากไม่มีการเปลี่ยนแปลง ตัวกรองภาพจะคงสภาพรูปภาพในพื้นที่นั้นไว้เช่นเดิม ซึ่งการปรับคุณภาพโดยการเน้นขอบของรูปภาพให้มีความชัดเจนยิ่งขึ้นคือการปรับค่าความเปรียบต่าง (contrast) ระหว่างสีที่แตกต่างกัน การเปลี่ยนแปลงอย่างรวดเร็วจากดำไปขาวทำให้ภาพคมชัดซึ่งต่างจากการจับเส้นขอบที่เห็นเฉพาะขอบและส่วนอื่นถูกลบออกไป ในการเปลี่ยนแปลงช้า ๆ จากดำเป็นเทาและขาวนั้นจะทำให้ภาพเบลอ
- 2) การกรองภาพแบบคอนโวลูชัน (Convolution Filtering) เป็นกระบวนการเฉลี่ยค่าของจุดภาพภายในหน้าต่างเคลื่อน (Sliding window) จากรูปที่ 2.1 หน้าต่างตัวกรองจะทำการคำนวณและนำผลลัพธ์มาใส่บนจุดภาพที่อยู่ตรงกลางของหน้าต่าง หลังจากการคำนวณเสร็จหนึ่งจุดภาพ หน้าต่างจะเลื่อนไปที่ละคอลัมน์และคำนวณจุดภาพที่อยู่ตรงกลางหน้าต่างตำแหน่งใหม่ เมื่อจบคอลัมน์ก็จะเลื่อนลงมาเริ่มต้นที่แถวใหม่ ทำจนหมดทั่วทั้งภาพ โดยเทคนิคที่ใช้ในการกรองภาพประกอบด้วยหลายเทคนิค เช่น เครื่องกรองผ่านความถี่สูง (high-pass filter) เป็นวิธีการกรองเอาองค์ประกอบที่มีความแปรปรวนในพื้นที่

สูงคือ พื้นที่ที่มีการเปลี่ยนแปลงของความถี่สูง รอยต่อของพื้นที่ที่แตกต่างกันมากหรือลักษณะที่เป็นเส้นขอบ



รูป 2.1 ขั้นตอนการกรองภาพด้วยการกรองแบบคอนโวลูชัน

2.2 การเรียนรู้ของเครื่อง (Machine Learning)

เป็นสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligent: AI) ที่พัฒนามาจากการศึกษาการรู้จำรูปแบบ (Pattern Recognition) โดยการเรียนรู้ของเครื่องเป็นการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้โดยอาศัยข้อมูลชุดตัวอย่างเพื่อตัดสินใจซึ่งแตกต่างจากการทำงานตามลำดับคำสั่งของโปรแกรม

- 1) การเรียนรู้ของเครื่องสามารถแบ่งออกได้เป็น 3 ประเภท ซึ่งถูกจำแนกโดยชุดข้อมูลฝึก (Training data) ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นการเรียนรู้ที่ผู้สอนต้องป้อนข้อมูลตัวอย่างและคำตอบของข้อมูลนั้น ๆ ซึ่งมีเป้าหมายเพื่อสร้างกฎที่สามารถเชื่อมโยงข้อมูลที่ถูกป้อนและผลลัพธ์ที่ออกมาได้
- 2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) การเรียนรู้ประเภทนี้ผู้สอนจะป้อนข้อมูลตัวอย่างเพียงอย่างเดียว ไม่มีการป้อนคำตอบของข้อมูลลงไป ซึ่งคอมพิวเตอร์จะเรียนรู้โดยการหาความคล้ายคลึงของข้อมูลที่ถูกป้อนและจัดกลุ่มข้อมูลได้อย่างเหมาะสม
- 3) การเรียนรู้แบบเสริมกำลัง (Reinforcement learning) การเรียนรู้ประเภทนี้ไม่จำเป็นต้องใช้ข้อมูล เนื่องจากเป็นการเรียนรู้แบบลองผิดลองถูกซึ่งจะมีการให้รางวัลและการทำโทษ (Reward and punishment) โดยจะต้องมีเป้าหมายใดเป้าหมายหนึ่งเป็นพิเศษ เช่น การเรียนรู้การขับยานพาหนะ หรือการเรียนรู้การเล่นเกม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การจำแนกข้อมูล (Data Classification)

การจำแนกข้อมูลเป็นกระบวนการสร้าง โมเดลหรือตัวคัดแยก (classifier) สำหรับจัดการข้อมูลให้อยู่ในกลุ่ม (class) ที่กำหนดมาให้จากกลุ่มตัวอย่างข้อมูลที่เรียกว่าชุดข้อมูลฝึก (Training data) ซึ่งในข้อมูล 1 ตัวอย่าง (instance) จะประกอบไปด้วยคุณสมบัติ (attribute) จำนวนมาก โดยจะมีบางคุณสมบัติที่เป็นตัวบ่งชี้กลุ่มของข้อมูล (Classifying attribute) ซึ่งจะสร้างความเชื่อมโยงระหว่างข้อมูลที่ถูกรับเข้ามาและคำตอบ

กระบวนการจำแนกข้อมูลประกอบไปด้วย 2 กระบวนการหลัก ได้แก่ การสร้างตัวจำแนกข้อมูล (Learning) และการทดสอบตัวจำแนกข้อมูล (Testing)

- 1) การสร้างตัวจำแนกข้อมูล (Learning) เป็นการนำชุดข้อมูลฝึก (Training data) มาหาความสัมพันธ์ระหว่างข้อมูลที่ถูกรับเข้ามาและคำตอบของข้อมูล (Class) เพื่อนำมาสร้างสมการหรือโมเดลสำหรับใช้คัดแยกหรือทำนายข้อมูล
- 2) การทดสอบตัวจำแนกข้อมูล (Testing) เป็นการนำตัวจำแนกข้อมูลที่ได้มาทดสอบและประเมินค่าความถูกต้องโดยการใช้ชุดข้อมูลทดสอบ (Testing data) ซึ่งเป็นข้อมูลที่ยังไม่เคยผ่านตัวจำแนกมาก่อน ทำให้ทราบถึงระดับความแม่นยำและสามารถประเมินได้ว่าสามารถนำมาใช้จริงได้หรือไม่ (การทดสอบด้วยชุดข้อมูลฝึกอาจก่อให้เกิดปัญหา overfitting ซึ่งเกิดจากการที่ตัวจำแนกข้อมูลมีความเฉพาะกับชุดข้อมูลฝึกหรือข้อมูลชุดใดชุดหนึ่งมากเกินไป)

2.4 การแบ่งกลุ่มข้อมูล (Clustering)

เป็นการจำแนกข้อมูลออกเป็นกลุ่มย่อย ๆ (cluster) ตั้งแต่ 2 กลุ่มขึ้นไป ซึ่งข้อมูลที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่เหมือน (similarity) หรือคล้ายกัน (proximity) โดยคำนวณจากการวัดระยะระหว่างเวกเตอร์ของข้อมูลโดยใช้การวัดระยะแบบต่าง ๆ เช่น การวัดระยะแบบยูคลิด (Euclidean distance) การวัดระยะแบบแมนฮัตตัน (Manhattan distance) หรือการวัดระยะแบบเชบิเชฟ (Chebychev distance) ซึ่งการแบ่งกลุ่มจะใช้ข้อมูลทั้งหมดในการหาจุดเด่นของแต่ละกลุ่มออกมาให้เห็นอย่างชัดเจนเพื่อให้ได้มาซึ่งลักษณะเฉพาะของกลุ่มนั้น ๆ

ขั้นตอนวิธีในการแบ่งกลุ่มข้อมูลโดยทั่วไปจะแบ่งได้เป็น 2 ประเภท

- 1) การแบ่งแบบลำดับขั้น (Hierarchical) เป็นการแบ่งกลุ่มจากกลุ่มย่อยที่ถูกแบ่งไว้ก่อนหน้านั้นซ้ำหลาย ๆ ครั้ง ซึ่งแบ่งออกเป็น 2 ลักษณะคือ แบบล่างขึ้นบน (Bottom-up) เป็นการแบ่งแบบรวมกลุ่มจากกลุ่มย่อยให้ใหญ่ขึ้นไปเรื่อย ๆ โดยเริ่มจากกลุ่มเล็กสุดคือในแต่ละกลุ่มมีข้อมูลเพียงตัวเดียว และแบบบนลงล่าง (Top-down) เป็นการแบ่งกลุ่มจากกลุ่มใหญ่ให้ย่อยลงไปเรื่อย ๆ โดยเริ่มจากกลุ่มใหญ่สุด
- 2) การแบ่งแบบตัดส่วน (Partitional Clustering) จะทำการแบ่งส่วนปริภูมิ (space) เพียงแค่ครั้งเดียวเท่านั้น

2.5 การคัดเลือกคุณสมบัติของข้อมูล (Feature Selection)

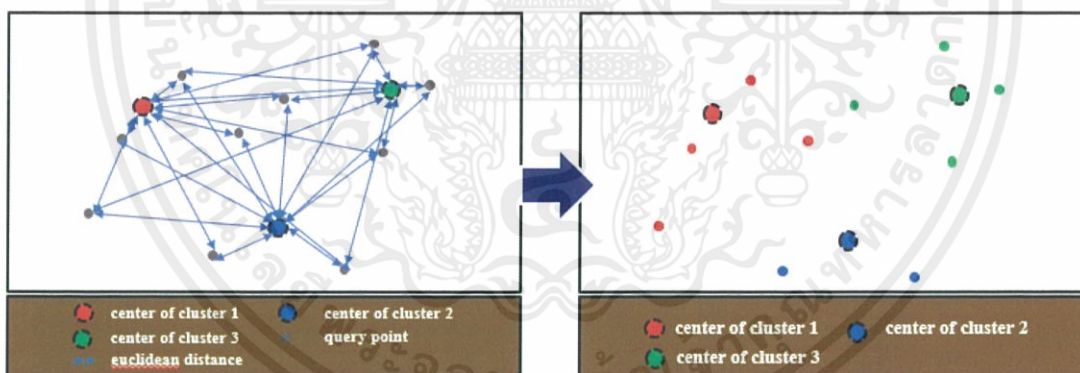
การคัดเลือกคุณสมบัติของข้อมูลหรือในอีกชื่อหนึ่งคือการคัดเลือกตัวแปร (Variable Selection) เป็นกระบวนการคัดเลือกสับเซตของคุณสมบัติที่มีความเกี่ยวข้องกันเพื่อใช้ในการสร้างโมเดลสำหรับทำนาย โดยการคัดเลือกคุณสมบัติของข้อมูลมีวัตถุประสงค์เพื่อลดความซับซ้อนของโมเดลเพื่อใช้ในการทำนาย ลดระยะเวลากระบวนการสร้างโมเดล และเพิ่มประสิทธิภาพการทำ generalization ของโมเดลเพื่อลดปัญหาการเกิด overfitting หรือ โมเดลมีความจำเพาะต่อข้อมูลชุดใดชุดหนึ่งเกินไป

แนวคิดของกระบวนการคัดเลือกคุณสมบัติของข้อมูลเกิดจากการที่ข้อมูลมีคุณลักษณะ (feature) ที่มากเกินไป มีความซ้ำซ้อนและไม่มีความเกี่ยวข้องกัน ซึ่งสามารถกำจัดคุณสมบัติเหล่านี้ได้โดยไม่ส่งผลต่อภาพรวมของข้อมูลหรือยังคงลักษณะเดิมของข้อมูลไว้ได้โดยไม่เปลี่ยนแปลงมากนัก โดยการคัดเลือกคุณสมบัติของข้อมูลจะถูกใช้กับโดเมนที่มีคุณลักษณะเป็นจำนวนมาก แต่มีจุดข้อมูล (Data point) ในจำนวนน้อย

2.6 การค้นหาเพื่อนบ้านใกล้ที่สุด (Nearest Neighbor Search)

การค้นหาเพื่อนบ้านใกล้ที่สุด (Nearest Neighbor Search: NSS) หรือการค้นหาความคล้าย (Similarity Search) เป็นวิธีการแก้ปัญหาสำหรับการหาจุดที่ใกล้เคียงที่สุดในปริภูมิ โดยความคล้าย จะถูกกล่าวถึงในแง่ของฟังก์ชันความแตกต่าง (Dissimilarity function) ยิ่งวัดความใกล้เคียงกัน น้อย ค่าที่ได้จากฟังก์ชันจะเพิ่มมากขึ้น ซึ่งปัญหาของ NSS จะถูกกำหนดโดยให้ S เป็นเซตของจุด ซึ่งอยู่ในปริภูมิ M และจุด q ซึ่งเป็นจุดที่ต้องการเปรียบเทียบ (Query point) โดยที่ q เป็นสมาชิกของ M

โดยส่วนใหญ่แล้วปริภูมิ M จะเป็นปริภูมิเมตริกซ์ (Metric space) และความแตกต่าง (Dissimilarity) จะถูกกล่าวถึงในรูปของระยะเมตริกซ์ (Distance metric) ซึ่งนอกจากนี้ปริภูมิ M ยังถูกกล่าวถึงในรูปของมิติใด ๆ บนปริภูมิเวกเตอร์ซึ่งความแตกต่างจะถูกวัดโดยใช้ระยะทางยูคลิด (Euclidean distance) ระยะทางแมนฮัตตัน (Manhattan distance) หรือระยะเมตริกซ์อื่น โดยอาศัยหลักการที่ว่า sample ที่มีลักษณะใกล้เคียงกัน จะมีระยะใกล้กัน โดยเราจะหาระยะโดยใช้ Euclidean distance ซึ่งเป็นระยะระหว่างจุด 2 จุด ยิ่ง Euclidean distance น้อย ยิ่งมีลักษณะใกล้เคียงกัน จากรูป 2.2 โดยโมเดลนี้จะหาระยะระหว่าง query point ซึ่งเป็น input กับศูนย์กลางของ cluster ต่าง ๆ ถ้าอยู่ใกล้ศูนย์กลางของ cluster ใด ก็จะถูกจัดให้อยู่ใน cluster นั้น



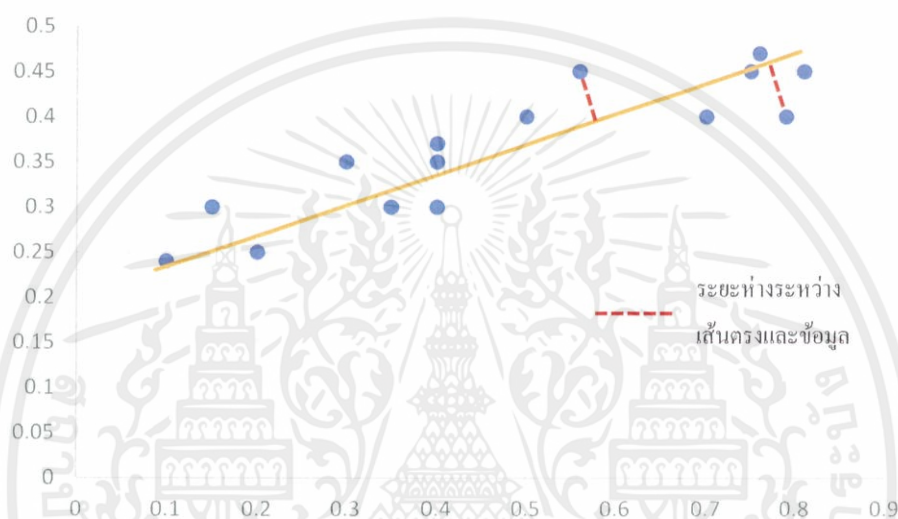
รูป 2.2 กระบวนการของอัลกอริทึม Nearest Neighbor Search

2.7 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis)

Linear Regression Analysis เป็นการวิเคราะห์เพื่อหาสมการความสัมพันธ์ระหว่างตัวแปร ตั้งแต่ 2 ตัวขึ้นไป โดยแบ่งประเภทตัวแปรออกเป็น 2 ประเภท คือ ตัวแปรอิสระหรือตัวแปรที่รู้จักค่า (Predictor) และตัวแปรตามหรือตัวแปรที่ไม่รู้จักค่า (Response) ซึ่งจะทำให้สามารถทราบขนาดความสัมพันธ์ระหว่างตัวแปรอิสระที่มีต่อตัวแปรตาม โดยการวิเคราะห์การถดถอยแบบเชิงเส้น

ตัวแปรอิสระส่วนใหญ่จะเป็นตัวแปรเชิงปริมาณ ในส่วนของตัวแปรตามจะต้องเป็นตัวแปรเชิงปริมาณเท่านั้น

ตัวอย่างของการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (รูปที่ 2.3) ซึ่งมีเพียงตัวแปรตาม 1 ตัว และตัวแปรอิสระ 1 ตัว จะนำข้อมูลจากตัวแปรทั้ง 2 มาวิเคราะห์หาความสัมพันธ์โดยใช้แผนภาพเส้นตรงและหาเส้นตรงที่ดีที่สุดเพื่อเป็นตัวแทนรูปแบบความสัมพันธ์ของตัวแปรที่ศึกษา โดยจะต้องมีผลรวมระยะห่างกำลังสองระหว่างเส้นตรงและทุก ๆ ข้อมูลมีค่าน้อยที่สุด (The Method of Least Squares)



รูป 2.3 การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

หลังจากได้เส้นตรงที่ดีที่สุดแล้ว เราจะได้สมการซึ่งประกอบไปด้วยค่าคงที่และสัมประสิทธิ์ หรือเรียกว่า สมการถดถอยเชิงเส้น เพื่อใช้ในการทำนายข้อมูล (Prediction)

2.8 การแบ่งกลุ่มด้วยวิธี Affinity Propagation

การแบ่งกลุ่มด้วยวิธี Affinity Propagation เป็นการแบ่งกลุ่มข้อมูลแบบ Unsupervised Learning โดยเป็นอัลกอริทึมที่ไม่จำเป็นต้องกำหนดกลุ่ม เช่น อัลกอริทึมจำพวก K-center clustering อีกทั้งยังมีอัตราการผิดพลาดน้อยเมื่อเทียบกับอัลกอริทึมอื่น ๆ หลักการทำงานของอัลกอริทึมนี้ คือ การมองทุกตัวอย่าง (sample) ให้เป็น โหนด (node) ที่มีการเชื่อมต่อ (edge) ถึงกัน โดยแต่ละโหนดจะส่งข้อมูลหากัน เพื่อหาโหนดที่มีลักษณะคล้ายกัน ซึ่งจะใช้โหนดที่มีลักษณะคล้ายเป็นตัวแทนโหนดใดที่ถูกอ้างอิงจากโหนดอื่นมากที่สุด โหนดนั้นจะเป็นจุดศูนย์กลางของกลุ่มนั้น ๆ

2.9 โครงข่ายประสาทเทียม OverFeat

OverFeat เป็นโครงข่ายประสาทเทียมแบบลึกชนิด Convolution ขนาด 8 ชั้น ซึ่งมีการเรียนรู้เป็นรูปภาพ (image-based) ถูกพัฒนาโดย Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus และ Yann LeCun ซึ่งเป็นโครงข่ายประสาทเทียมที่ชนะเลิศการแข่งขัน ImageNet ในปี ค.ศ. 2013 ถูกสร้างด้วยรูปภาพซึ่งเป็นชุดข้อมูลฝึก (Training data) จำนวน 1.2 ล้านรูปและมีอัตราความผิดพลาดน้อยที่สุด 1 ใน 5 ของการแข่งขันที่ 14.20 %

OverFeat สามารถทำการจำแนก (Classification) ประเภทของวัตถุที่อยู่ในภาพรวมถึงระบุตำแหน่งของวัตถุบนภาพ (Localization) นอกจากนี้ยังสามารถทำการดึงคุณสมบัติ (Feature Extraction) ออกมาจากภาพได้ โดยจะให้คุณสมบัติทั้งหมด 4096 คุณสมบัติต่อ 1 ภาพ และภาพที่จะผ่านการดึงคุณสมบัติจะต้องมีความละเอียดไม่ต่ำกว่า 231x231 พิกเซล

2.10 การหาระยะทางแบบยูคลิด (Euclidean Distance)

การวัดความห่างระหว่างจุดสองจุด เป็นวิธีการหนึ่งในการคัดเลือกวัตถุที่มีความใกล้เคียงกันมากที่สุด โดยใช้ค่าของความห่างระหว่างจุดมาเป็นตัวตัดสินว่าวัตถุใดมีความใกล้เคียงมากที่สุด ซึ่งวิธีนี้สามารถหาความห่างในแนวเส้นตรงซึ่งสามารถพิสูจน์หาค่าได้ด้วยทฤษฎีบทพีทาโกรัส โดยเป็นการวัดระยะห่างในปริภูมิแบบยูคลิด คือ การวัดแบบไม่มีความโค้งและไม่สามารถทำให้โค้งงอได้ (ปริภูมิแบบยูคลิด) ด้วยเหตุนี้จึงเรียกอวกาศที่แบบนี้เป็นว่า แบบยูคลิด

นิยาม คือ ระยะทางระหว่างจุด $P = (p_1, p_2, \dots, p_n)$ และ $Q = (q_1, q_2, \dots, q_n)$ ในระบบพิกัดคาร์ทีเซียนเป็นจุดสองจุดบนปริภูมิยูคลิด n มิติ ระยะทางระหว่างจุด p กับ q คำนวณได้จากสมการ ดังนี้

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}. \quad (1.1)$$

ค่าประจำแบบยูคลิด คือระยะทางจากจุดหนึ่งจุด p ไปยังจุดกำเนิด $(0, 0, \dots, 0)$ บนปริภูมิยูคลิดซึ่งสมการตัวหลังเกี่ยวข้องกับผลคูณจุด เป็นขนาดของเวกเตอร์ p จากจุดกำเนิด ระยะทางแบบยูคลิดจึงอาจนิยามได้อีกแบบหนึ่งดังนี้

2.10.1. ระยะทางในขนาด 1 มิติ

หาระยะทางระหว่าง 2 จุดในขนาด 1 มิติ $P = (p_x)$ และ $Q = (q_x)$

หาระยะทางได้โดย $\sqrt{(p_x - q_x)^2} = |p_x - q_x|$.

2.10.2. ระยะทางในขนาด 2 มิติ

หาระยะทางระหว่าง 2 จุดในขนาด 2 มิติ $P = (p_x, p_y)$ และ $Q = (q_x, q_y)$

หาระยะทางได้โดย $\sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$.

2.10.3. ระยะทางในขนาด 3 มิติ

หาระยะทางระหว่าง 2 จุดในขนาด 3 มิติ $P = (p_x, p_y, p_z)$ และ $Q = (q_x, q_y, q_z)$ หาระยะทางได้โดย $\sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2}$.

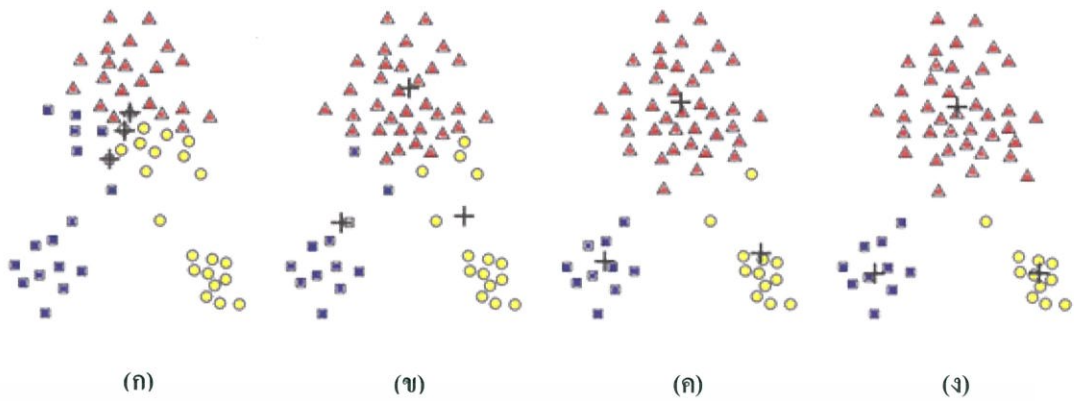
2.10.4. ระยะทางในขนาดหลายมิติ

หาระยะทางระหว่าง 2 จุดในขนาดหลายมิติ $P = (p_1, p_2, \dots, p_n)$ และ $Q = (q_1, q_2, \dots, q_n)$ หาระยะทางได้โดย $\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$.

2.11 การแบ่งกลุ่มข้อมูลแบบเคมีน (K-Means Clustering)

การแบ่งกลุ่มข้อมูลเป็นวิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) โดยอัลกอริทึมของ K-Means เริ่มแรกจะทำการจัดกลุ่มโดยการกำหนดจำนวนกลุ่ม (K) ที่ต้องการ และกำหนดจุดศูนย์กลางเริ่มต้นของแต่ละกลุ่ม ซึ่งการกำหนดตำแหน่งจุดศูนย์กลางเริ่มต้นที่แตกต่างกันทำให้ได้ผลลัพธ์สุดท้ายแตกต่างกันด้วย ดังนั้นจึงต้องมีการกำหนดจุดศูนย์กลางให้ห่างจากจุดศูนย์กลางอื่น ๆ ให้มากที่สุด ขั้นตอนต่อไปคือสร้างกลุ่มข้อมูลและความสัมพันธ์กับจุดศูนย์กลางที่ใกล้มากที่สุด โดยแต่ละจุดจะถูกกำหนดให้เป็นกลุ่มเดียวกับจุดศูนย์กลางที่ใกล้ที่สุดจนครบหมดทุกจุด และคำนวณจุดศูนย์กลางใหม่โดยการหาค่าเฉลี่ยทุกจุดที่อยู่ในกลุ่ม หากจุดศูนย์กลางในแต่ละกลุ่มถูกเปลี่ยนตำแหน่ง จะได้จุดที่มีความสัมพันธ์กับกลุ่มใหม่และใกล้กับจุดศูนย์กลางใหม่ ทำซ้ำแบบนี้ไปเรื่อย ๆ จะสังเกตเห็นว่าผลลัพธ์จากการทำซ้ำแบบนี้ทำให้จุดศูนย์กลางเปลี่ยนตำแหน่งทุกรอบ จนกระทั่งจุดศูนย์กลางจำนวน K จุด ไม่มีการเปลี่ยนแปลงจึงจะสิ้นสุดกระบวนการ สามารถสรุปขั้นตอนเริ่มต้นได้ดังนี้

- 1) กำหนดจำนวนและจุดศูนย์กลางเริ่มต้นจำนวน K จุด
- 2) นำจุดทั้งหมดจัดเข้ากลุ่มที่มีจุดศูนย์กลางที่อยู่ใกล้จุดนั้นมากที่สุด โดยคำนวณจากการวัดระยะห่างระหว่างจุดที่น้อยที่สุด
- 3) คำนวณจุดศูนย์กลาง K จุดใหม่ โดยหาจากค่าเฉลี่ยทุกจุดที่อยู่ในกลุ่ม
- 4) ทำซ้ำในข้อ 2. จนกระทั่งจุดศูนย์กลางไม่เปลี่ยนแปลง



รูป 2.14 ขั้นตอนการทำงานของ k-mean

- (ก) การกำหนดจำนวนกลุ่มและจุดศูนย์กลางเริ่มต้น
- (ข) จัดจุดทั้งหมดให้เข้ากลุ่ม
- (ค) ปรับจุดศูนย์กลางใหม่และจัดให้เข้ากลุ่ม
- (ง) ผลลัพธ์สุดท้าย

รูป 2.14 (ก) เป็นการจัดกลุ่มในขั้นตอนแรกโดยที่กำหนดจำนวนกลุ่มและจุดศูนย์กลางเริ่มต้นจำนวน 3 กลุ่ม โดยใช้สัญลักษณ์ + แทนจุดศูนย์กลางของแต่ละกลุ่ม จากนั้นวัตถุประสงค์จะกำหนดให้เข้ากลุ่มที่มีจุดศูนย์กลางอยู่ใกล้ตัวจุดนั้นมากที่สุดดังรูป 2.14 (ข) จุดศูนย์กลางมีการเปลี่ยนแปลงเป็นจุดศูนย์กลางใหม่และจัดจุดทั้งหมดให้เข้ากลุ่มที่มีจุดศูนย์กลางอยู่ใกล้กับตัวจุดนั้นมากที่สุดดังรูป 2.14 (ค) ทำซ้ำเช่นนี้ไปเรื่อย ๆ จนกระทั่งจุดศูนย์กลางไม่เปลี่ยนแปลงจึงจะได้ผลลัพธ์สุดท้ายดังรูป 2.14 (ง)

หลังจากนั้น เมื่อมีข้อมูลใหม่เข้ามาอัลกอริทึมนี้จะทำการวัดระยะห่างของข้อมูลนั้นกับจุดศูนย์กลางของแต่ละกลุ่ม ข้อมูลจะถูกจัดเข้าไปอยู่ในกลุ่มที่มีระยะห่างน้อยที่สุดและหาค่าเฉลี่ยของจุดศูนย์กลางใหม่

2.12 การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis)

การวิเคราะห์องค์ประกอบหลักเป็นวิธีหนึ่งในการสกัดคุณสมบัติ (Factor Extraction) ที่ช่วยลดจำนวนตัวแปรและสามารถอธิบายความแปรปรวนของข้อมูลได้อย่างครบถ้วน การวิเคราะห์องค์ประกอบหลัก (PCA) เป็นวิธีการทางสถิติที่มีความใกล้เคียงกับการวิเคราะห์ปัจจัย (Factor Analysis) ที่มีวัตถุประสงค์หลักเพื่อลดจำนวนตัวแปร (data reduction) และสร้างปัจจัย (Factor) จากตัวแปร เพื่อศึกษาแบบแผนและโครงสร้างความสัมพันธ์ของตัวแปรเพื่อพัฒนาและตรวจสอบความเที่ยงตรงของโมเดลการวัดที่สร้างขึ้น ซึ่งเป็นวิธีที่ใช้สกัดองค์ประกอบที่มีความร่วมที่เหมือนกันจากตัวแปรจำนวนมาก

2.13 เทคโนโลยีที่ใช้

2.13.1. SQL Database



รูป 2.4 สัญลักษณ์ของ MySQL

เป็นเทคโนโลยีการจัดการฐานข้อมูลชนิดหนึ่งที่ใช้ในการเรียกและเก็บข้อมูลในเชิงความสัมพันธ์ (relational database management system: RDBMS) ผ่านภาษา SQL ซึ่งเป็นภาษาที่เข้าใจง่ายไม่ซับซ้อน แต่มีประสิทธิภาพในการทำงานสูง จัดการง่าย รวมถึงมีระยะเวลาในการตอบสนองต่อคำสั่งที่รวดเร็ว จึงทำให้เป็นที่นิยมและได้รับการรับรองให้เป็นมาตรฐานหนึ่งของสถาบันมาตรฐานสหรัฐอเมริกา (American National Standards Institute - ANSI) และองค์การมาตรฐานสากล (International Standards Organization - ISO) ซึ่งในโครงการนี้จะใช้ MySQL ซึ่งเป็น SQL Database ชนิดหนึ่งที่เป็นที่นิยมกันมาก เนื่องจากเป็นฟรีแวร์ทางด้านฐานข้อมูลที่มีประสิทธิภาพสูง ผู้ใช้ต่างยอมรับในความรวดเร็วและการรองรับจำนวนผู้ใช้และขนาดของข้อมูลจำนวนมาก

2.13.2. Django Framework



รูป 2.5 สัญลักษณ์ของ Django

เป็นชุดเครื่องมือ Framework สำหรับ การพัฒนาเว็บไซต์ด้วยภาษา Python ซึ่งมี Framework สำหรับการเขียนเว็บไซต์ด้วยภาษา Python จำนวนมาก โดย Django Framework เป็นหนึ่งใน Framework สำหรับการพัฒนาเว็บไซต์ด้วยภาษา Python โดยปัจจุบันภาษา Python นั้นได้รับความนิยมเพิ่มมากขึ้น ดังนั้นในโครงการนี้เลือกใช้ Django

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.14 เครื่องมือที่ใช้

2.14.1. ภาษาที่ใช้ในการพัฒนา

2.14.1.1. Python



รูป 2.6 สัญลักษณ์ของภาษา Python

ภาษา Python เป็นภาษา script ที่มีความเรียบง่ายต่อการเขียนโปรแกรมเพื่อการพัฒนา มีประสิทธิภาพสูงและใช้เวลาในการ compile น้อย อีกทั้งยังมี library ที่สนับสนุนการคำนวณด้านวิทยาศาสตร์มากมาย เหมาะที่จะใช้ในการทำ machine learning

2.14.1.2. Java



รูป 2.7 สัญลักษณ์ของภาษา Java

ภาษา Java เป็นภาษาที่พัฒนามาแทนที่ภาษา C++ ภาษา Java จึงมีรูปแบบคล้ายกับภาษา C/C++ โดยภาษา Java นั้นเป็นภาษาเชิงวัตถุ (Object Oriented Programming)

2.14.2. Library ที่ใช้ในการพัฒนา

Scikit-learning



รูป 2.8 สัญลักษณ์ของ Scikit-learning

Scikit-learning เป็น library ของภาษา python สำหรับใช้ในการทำ machine learning, data mining และ data analysis ซึ่งมีการเตรียม training set, testing set และ example เพื่อเป็นตัวอย่าง นอกจากนี้ยังเป็น library ที่มีความง่ายในการเรียกใช้งานฟังก์ชันต่าง ๆ

2.14.3. IDE ที่ใช้ในการพัฒนา

2.14.3.1. Android Studio



รูป 2.9 สัญลักษณ์ของ Android Studio

Android Studio เป็น IDE สำหรับพัฒนาแอปพลิเคชันบนระบบปฏิบัติการ Android พัฒนาโดย JetBrains นอกจากนี้ยังมี emulator ซึ่งทำให้สามารถทดสอบแอปพลิเคชันที่ ถูกพัฒนาขึ้นมากับอุปกรณ์ที่มีหน้าจอต่างขนาดกัน

2.14.3.2. JetBrains PyCharm



รูป 2.10 สัญลักษณ์ของ JetBrains PyCharm

PyCharm เป็น IDE ถูกพัฒนาโดย JetBrains มีความสามารถในการช่วยอำนวยความสะดวกการใช้งาน framework ภาษา python เช่น Django Flask หรือ Google App Engine เป็นต้น

บทที่ 3

การออกแบบและพัฒนา

3.1 ความต้องการของระบบและฟังก์ชันการทำงาน

3.1.1 รายละเอียดข้อมูลที่ระบบต้องการ (Input Specification)

- 1) ลักษณะและรูปภาพไบหน้าตรงของสุนัขรวมถึงพิกัดเจ้าของสุนัข
- 2) ลักษณะและรูปภาพไบหน้าตรงของสุนัขที่ถูกพบรวมถึงพิกัดที่พบและข้อมูลติดต่อผู้พบสุนัข

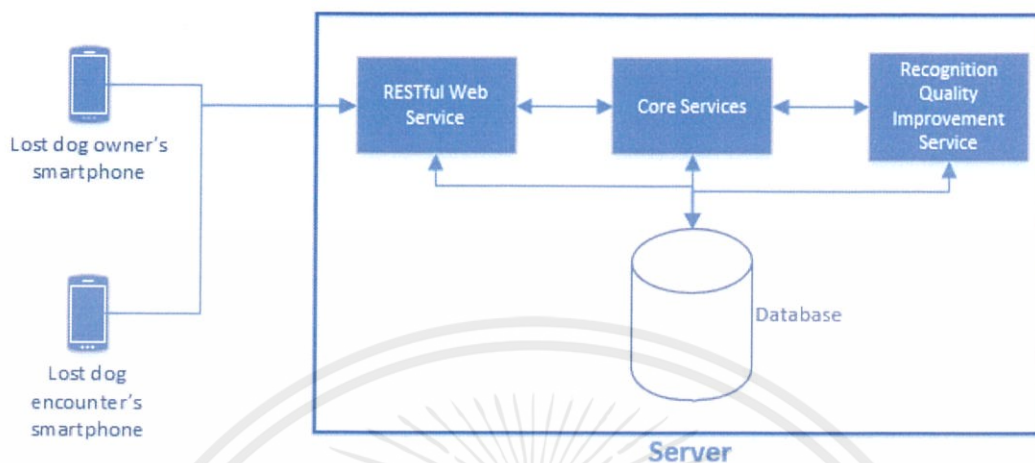
3.1.2 รายละเอียดผลลัพธ์ของระบบ (Output Specification)

รูปภาพไบหน้าสุนัขที่มีความใกล้เคียงกับรูปภาพไบหน้าสุนัขที่ถูกป้อนเข้าระบบ รวมถึงพิกัดที่พบและข้อมูลติดต่อผู้พบสุนัข

3.1.3 ฟังก์ชันการทำงานของระบบ (Functional Specification)

- 1) เจ้าของสุนัขสูญหายสามารถค้นหารูป ลักษณะและรายละเอียดของสุนัขสูญหายโดยใช้รูปถ่ายหน้าตรงของสุนัขที่มีอยู่หรือรูปถ่ายที่ได้บันทึกลงในระบบก่อนหน้านั้น โดยระบบจะค้นหาข้อมูลและรูปภาพสุนัขที่มีลักษณะใกล้เคียงกับรูปภาพสุนัขที่ถูกป้อนเข้ามาในระบบมากที่สุด
- 2) เจ้าของหรือผู้พบเจอสุนัขสูญหายสามารถสร้างประกาศหาหรือประกาศพบสุนัขสูญหายได้
- 3) ระบบแจ้งเตือนอัตโนมัติเมื่อมีผู้พบสุนัขที่มีลักษณะใกล้เคียงกับสุนัขที่หาย

3.2 โครงสร้างของระบบ



รูป 3.1 แผนผังโครงสร้างของระบบ

จากรูป 3.1 ผู้ใช้จะแบ่งเป็น 2 ประเภทคือ เจ้าของสุนัขสูญหาย (Lost Dog Owner) และผู้พบเห็นสุนัข (Lost Dog Encounterer) ในส่วนของระบบบนเซิร์ฟเวอร์จะประกอบไปด้วย 3 ระบบ

3.2.1 RESTful Web Service

Service สำหรับให้บริการการใช้งาน Web API ตามที่ผู้ใช้งานขอ เป็นตัวกลางในการเชื่อมต่อประสานระหว่างโทรศัพท์มือถือของผู้ใช้กับบริการหลักของระบบ เมื่อผู้ใช้งาน (ผ่าน URI) บริการจะทำหน้าที่ส่งต่อคำร้องไปยังบริการย่อยที่มีหน้าที่รับผิดชอบที่อยู่ในบริการหลักของระบบ (Core Service)

3.2.2 Core Services

Core Services เป็นบริการหลักของระบบ โดยประกอบไปด้วยระบบย่อย 4 ระบบ ได้แก่

3.2.2.1 ระบบจัดการบัญชีผู้ใช้ (User Management System)

เป็นระบบสำหรับจัดการข้อมูลส่วนตัวของผู้ใช้แต่ละคน

3.2.2.2 ระบบติดตามสุนัขที่อยู่ในการดูแล (Dog Tracking System)

เป็นระบบสำหรับเพิ่มข้อมูลสุนัขที่อยู่ในการดูแลของผู้ใช้งาน เช่น รูปใบหน้าสุนัข สายพันธุ์หรือลักษณะเพื่อใช้สำหรับติดตามในกรณีที่สุนัขสูญหาย

3.2.2.3 ระบบค้นหาสุนัขด้วยรูป (Dog Search Engine)

เป็นระบบสำหรับค้นหาสุนัขสูญหายโดยอาศัยรูปสุนัขที่ได้จากเจ้าของสุนัขสูญหายและรูปจากผู้พบเห็นสุนัขซึ่งอยู่ในฐานข้อมูล ประกอบไปด้วย 3 ส่วน

ได้แก่ ส่วนสกัด feature บนใบหน้าสุนัข (Feature Extractor) ส่วนระบุ class ของ feature (Classifier) และส่วนเทียบความเหมือนใบหน้าสุนัข (Similar Dog Face Comparator) โดยให้ผลลัพธ์เป็นรูปและข้อมูลของสุนัขที่มีความใกล้เคียงกับรูปสุนัขสูญหายที่ถูกป้อนเข้ามามากที่สุดจนถึงน้อยสุดตามลำดับ ผู้ใช้สามารถดูข้อมูลของสุนัขได้เช่นรูปทั้งหมด เวลาและพิกัดที่พบ

3.2.2.4 ระบบประกาศตามหาและประกาศพบสุนัขหาย (Lost&Found System)

เป็นระบบสำหรับประกาศเพื่อช่วยตามหาสุนัขสูญหาย โดยระบบจะแจ้งเตือนอัตโนมัติเมื่อผู้พบเจอใส่รูปสุนัขที่สูญหายเข้ามาในระบบและมีแนวโน้มหรือลักษณะคล้ายหรือเป็นตัวเดียวกัน ผู้พบเจอสามารถบันทึกรูปภาพสุนัขรวมถึงข้อมูลอื่น ๆ เช่น คำอธิบายลักษณะ หรือเวลาและพิกัดที่พบเจอ

3.2.2.5 ระบบปรับปรุงคุณภาพอัลกอริทึมสำหรับการค้นหา (Searching Algorithm Improvement System)

ระบบสำหรับปรับปรุงให้ผลการค้นหามีความใกล้เคียงกับรูปสุนัขที่ป้อนเข้ามาในระบบมากที่สุด และสามารถรองรับการรู้จำสุนัขพันธุ์ใหม่เพิ่มขึ้นหรือสุนัขที่มีลักษณะผสมได้ตลอดเวลา

3.3 การออกแบบอัลกอริทึมสำหรับคัดแยกและค้นหาใบหน้าสุนัข

ในการออกแบบอัลกอริทึมสำหรับค้นหาใบหน้าสุนัขจะต้องคำนึงถึงความคล้ายของใบหน้าสุนัขที่ป้อนเข้ามากับใบหน้าสุนัขที่มีในระบบ ยิ่งคล้ายกันมากเท่าใด โอกาสที่เจ้าของจะพบเจอสุนัขยังมีมากขึ้น ระยะเวลาที่ใช้ในการทำงานของอัลกอริทึม หากไม่มีการออกแบบที่ดีพอจะทำให้ใช้ระยะเวลาในการประมวลผลตรงส่วนนี้เป็นเวลานาน เนื่องจากเวลาที่ถูกใช้ไปในส่วนอื่นมากพอเพียงแล้ว เช่น การส่งข้อมูลระหว่างผู้ใช้และเซิร์ฟเวอร์ การประมวลผลของเซิร์ฟเวอร์เอง การติดต่อระหว่างเซิร์ฟเวอร์และบริการต่าง ๆ อาจไม่เป็นที่พอใจของผู้ใช้ จึงมีความจำเป็นที่จะต้องออกแบบอัลกอริทึมให้ใช้เวลาในการทำงานน้อยที่สุดเพื่อลดเวลาการทำงานโดยรวมของระบบ นอกจากนี้อัลกอริทึมต้องสามารถรองรับสายพันธุ์ใหม่ ๆ หรือลักษณะใหม่ ๆ ในกรณีที่เป็นสุนัขพันธุ์ผสมได้ตลอดเวลา

การออกแบบจะแบ่งออกเป็น 3 ส่วน คือ อัลกอริทึมสำหรับดึงคุณสมบัติจากรูปใบหน้าสุนัข อัลกอริทึมสำหรับคัดแยกใบหน้าสุนัขตามกลุ่มที่มีลักษณะใบหน้าคล้ายกันและอัลกอริทึมสำหรับค้นหาใบหน้าสุนัขที่มีลักษณะคล้ายที่สุด

3.3.1 อัลกอริทึมสำหรับสร้างตัวคัดเลือกคุณสมบัติจากรูปใบหน้าสุนัข

OverFeat สามารถดึงคุณสมบัติจากรูปภาพได้ โดยจะให้คุณสมบัติทั้งหมด 4,096 คุณสมบัติต่อ 1 รูปภาพ ซึ่งจำนวนที่มากดังกล่าวอาจก่อให้เกิดปัญหามิติของข้อมูลหรือ Curse of Dimensionality เกิดจากการที่มีมิติข้อมูลไม่ได้ไปในทิศทางเดียวกัน ข้อมูลเกิดการกระจาย บางจุดอาจจะไม่มีข้อมูลอยู่เลย ส่งผลโดยรวมต่อการนำข้อมูลไปใช้ โดยเฉพาะการแบ่งกลุ่มด้วย K-Means ซึ่งให้นำหนักทุก ๆ คุณสมบัติเท่ากัน อาจก่อให้เกิดปัญหาการแบ่งกลุ่มโดยข้อมูลไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์ต่ำ จึงต้องทำการลดคุณสมบัติที่ไม่สำคัญออกไปโดยการคัดเลือกคุณสมบัติ ซึ่งสามารถใช้การวิเคราะห์การถดถอยเชิงเส้นหาค่าน้ำหนักของแต่ละคุณสมบัติได้ และนำค่าดังกล่าวมาคัดคุณสมบัติที่ไม่มีน้ำหนักหรือมีน้ำหนักต่ำออก

กระบวนการคัดเลือกคุณสมบัติสามารถแบ่งออกได้เป็น 8 ขั้นตอน

- 1) การเตรียมรูปภาพใบหน้าสุนัข ซึ่งเป็นรูปใบหน้าสุนัขหน้าตรง โดยยอมให้มีการเอียงของใบหน้าได้แต่ต้องมียอดศีรษะประกอบบนใบหน้าครบ ได้แก่ ตาทั้งสองข้าง หมุก ปาก ซึ่งในที่นี้จะเก็บรูปภาพใบหน้าสุนัข 30 ตัว ที่มีลักษณะแตกต่างกัน จำนวน 30 สายพันธุ์



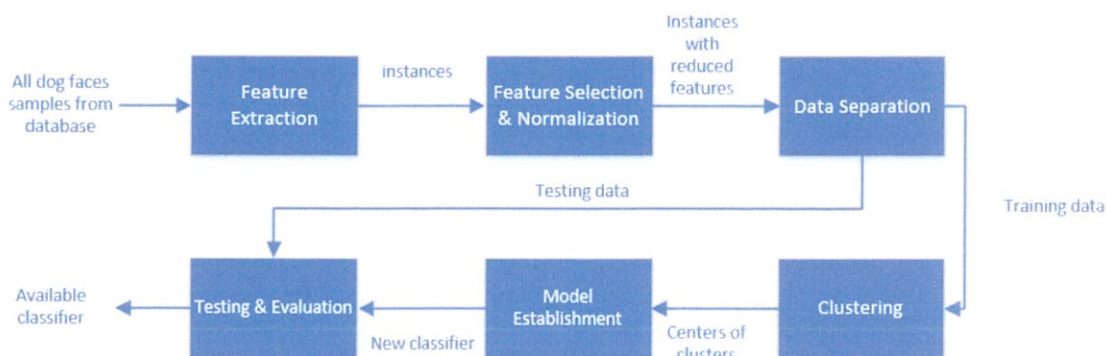
รูป 3.2 การจัดเตรียมรูปใบหน้าสุนัข

- 2) ดึงคุณสมบัติจากรูปภาพใบหน้าสุนัข โดยใช้ OverFeat ในการดึงคุณสมบัติ 4096 คุณสมบัติต่อ 1 รูปภาพ จากนั้นนำข้อมูลทั้งหมดเก็บลงฐานข้อมูล โดยเก็บแยกของสุนัขแต่ละตัว
- 3) จับคู่และรวมคุณสมบัติ โดยการจับคู่คุณสมบัติของรูปใบหน้าสุนัข 2 รูป รูปละ 4096 คุณสมบัติ รวมเป็น 8192 คุณสมบัติ การจับคู่ในขั้นตอนนี้จะแบ่งออกเป็น 2 แบบคือ การจับคู่ที่ถูกต้อง เป็นการจับคู่คุณสมบัติของสุนัขตัวเดียวกัน และการจับคู่ที่ไม่ถูกต้อง เป็นการจับคู่คุณสมบัติของสุนัขคนละสายพันธุ์ โดยการจับคู่ที่ถูกต้องจะกำหนด class เป็น 1 ส่วนการจับคู่ที่ไม่ถูกต้องจะกำหนด class เป็น 0
- 4) สุ่มการจับคู่ โดยรักษาอัตราส่วนระหว่างการจับคู่ที่ถูกต้องและการจับคู่ที่ไม่ถูกต้องให้เท่ากัน เพื่อลดความผิดพลาดของการสร้างโมเดล เนื่องจากเน้นที่การนำค่าน้ำหนักที่ได้จากโมเดลมาใช้เป็นสำคัญ
- 5) นำคุณสมบัติที่ถูกจับคู่ทั้งหมดมาสร้างโมเดลโดยใช้การวิเคราะห์การถดถอยเชิงเส้น ซึ่งเราจะได้ค่าน้ำหนักหรือสัมประสิทธิ์ (coefficient) ของคุณสมบัติแต่ละตัว
- 6) นำคุณสมบัติทั้งหมดมาผ่านการกรอง ค่าน้ำหนักของคุณสมบัติใดไม่ถึงเกณฑ์ที่กำหนดจะถูกคัดออก คงเหลือเฉพาะคุณสมบัติที่สำคัญเท่านั้น
- 7) นำคุณสมบัติของสุนัขทั้งหมดมาสร้างโมเดลโดยใช้อัลกอริทึมการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) โดยกำหนดให้อัตราการรักษาองค์ประกอบอยู่ที่ร้อยละ 90
- 8) นำคุณสมบัติของสุนัขทั้งหมดมาผ่านการแปลงปริภูมิ (Dimensional Transformation) เพื่อลดปริภูมิ ส่งผลโดยตรงต่อการลดจำนวนคุณสมบัติ

3.3.2 อัลกอริทึมสำหรับคัดแยกใบหน้าสุนัขตามกลุ่มที่มีลักษณะใบหน้าคล้ายกัน

ในขั้นตอนนี้ จะนำข้อมูลของรูปใบหน้าสุนัขทั้งหมดที่ผ่านขั้นตอนการดึงคุณสมบัติ และขั้นตอนการคัดเลือกคุณสมบัติแล้วมาสร้างตัวคัดแยกขึ้นมาเพื่อใช้คัดแยกรูปใบหน้าสุนัขให้อยู่ในกลุ่มที่มีลักษณะใบหน้าที่คล้ายกัน แบ่งออกเป็น 4 ขั้นตอน

- 1) การแบ่งกลุ่มด้วยอัลกอริทึม Affinity Propagation



รูป 3.3 ขั้นตอนการสร้างโมเดล

- 2) ขั้นตอนการแบ่งข้อมูล ในขั้นตอนนี้จะแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ชุดข้อมูลฝึกและชุดข้อมูลทดสอบ โดยจะนำชุดข้อมูลฝึกไปใช้ในกระบวนการสร้างโมเดล และนำชุดข้อมูลทดสอบไปใช้ในกระบวนการทดสอบและประเมินโมเดล เพื่อป้องกันการเกิดปัญหา overfitting โดยข้อมูลของสุนัข 1 ตัว จะถูกแบ่งออกเป็น 2 ส่วน ด้วยอัตราส่วนชุดฝึกและชุดทดสอบ 7:3 ส่วน
- 3) ขั้นตอนการสร้างโมเดลหรือตัวคิดแยก นำชุดข้อมูลฝึกมาสร้างโมเดล
- 4) ขั้นตอนทดสอบและประเมินโมเดล เป็นขั้นตอนในการนำชุดข้อมูลทดสอบซึ่งไม่เคยผ่านตัวคิดแยก นำมาป้อนและตรวจสอบผลลัพธ์ที่ได้ หากค่าความถูกต้องไม่ถึงเกณฑ์ จะต้องมีการปรับพารามิเตอร์ของโมเดลหรือปรับปรุงคุณภาพของข้อมูลเพื่อให้ได้ค่าความถูกต้องเพียงพอสำหรับการทำงานกับข้อมูลจริง

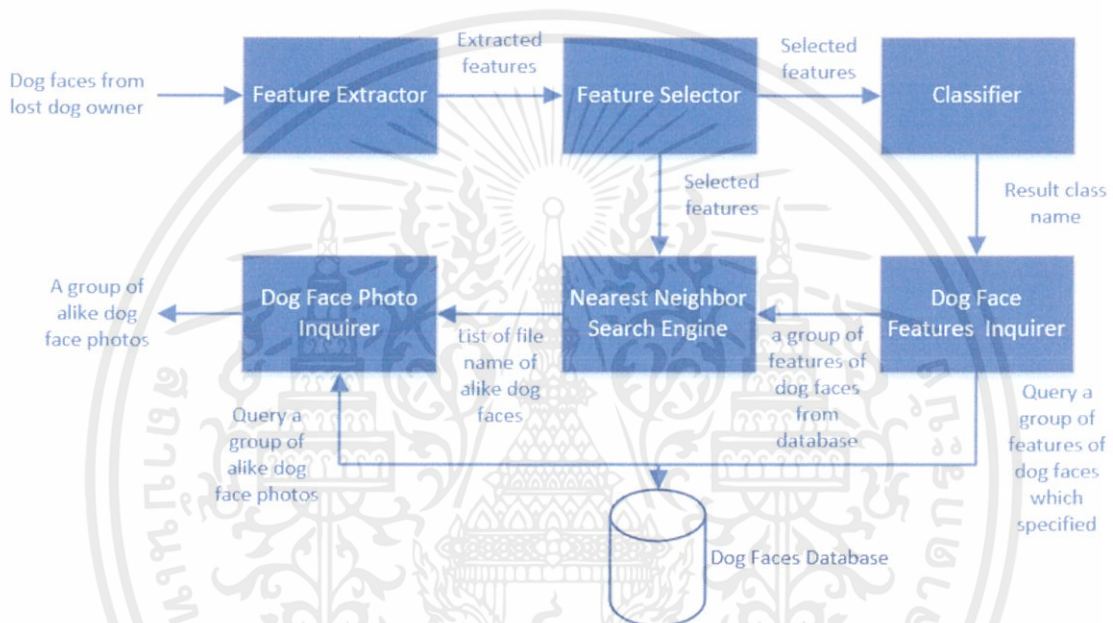
3.3.3 อัลกอริทึมสำหรับค้นหาใบหน้าสุนัขที่มีลักษณะคล้ายที่สุด

เราสามารถค้นหาใบหน้าสุนัขที่มีลักษณะคล้ายกันได้โดยอาศัยคุณสมบัติที่มีค่าต่างกันของแต่ละรูปใบหน้าสุนัข ด้วยหลักการคือยิ่งลักษณะใบหน้าสุนัขบนรูปภาพมีความคล้ายกันมากเท่าใด ความแตกต่างของคุณสมบัติก็ยิ่งน้อยลง ซึ่งสามารถดำเนินการได้ด้วยการวิธีการค้นหาเพื่อนบ้านใกล้เคียงหรือ Nearest Neighbor Search โดยใช้อัลกอริทึมการหาระยะทางยูคลิเดียนในการหาระยะบนปริภูมิระหว่างจุดที่ต้องการ (Query point) ซึ่งก็คือรูปใบหน้าสุนัขสูญหาย เทียบกับทุก ๆ จุดบนปริภูมิซึ่งก็คือใบหน้าสุนัขที่มีในระบบ โดยจะได้เป็นระยะออกมา ยิ่งระยะของจุดใด ๆ เทียบกับจุดที่ต้องการเข้าใกล้ศูนย์ ค่าความคล้ายก็จะเพิ่มขึ้นมากเท่านั้น แต่เนื่องจากข้อมูลของสุนัขในระบบมีจำนวนมากและเพิ่มขึ้นเรื่อย ๆ ตามจำนวนผู้ใช้ที่เพิ่มขึ้น จึงจำเป็นต้องแบ่งออกเป็นกลุ่มแยก ๆ ตามลักษณะความคล้ายแล้วจึงใช้ NSS ในการหารูปภาพใบหน้าสุนัขที่คล้ายกัน เพื่อลด

ระยะเวลาการทำงานของอัลกอริทึมลง เนื่องจากการเทียบทุก ๆ จุดข้อมูลจะทำให้ใช้เวลานาน

3.4 การทำงานของอัลกอริทึมค้นหาใบหน้าสุนัข

อัลกอริทึมจะรับข้อมูลจากผู้ใช้งานซึ่งเป็นรูปภาพใบหน้าสุนัขที่เน้นเฉพาะใบหน้าเพียงอย่างเดียวเท่านั้นเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด โดยอัลกอริทึมจะให้ผลลัพธ์เป็นรูปภาพสุนัขที่มีลักษณะใกล้เคียงที่สุดจำนวนรูปตามที่กำหนดเพื่อให้ผู้ใช้พิจารณาต่อไป



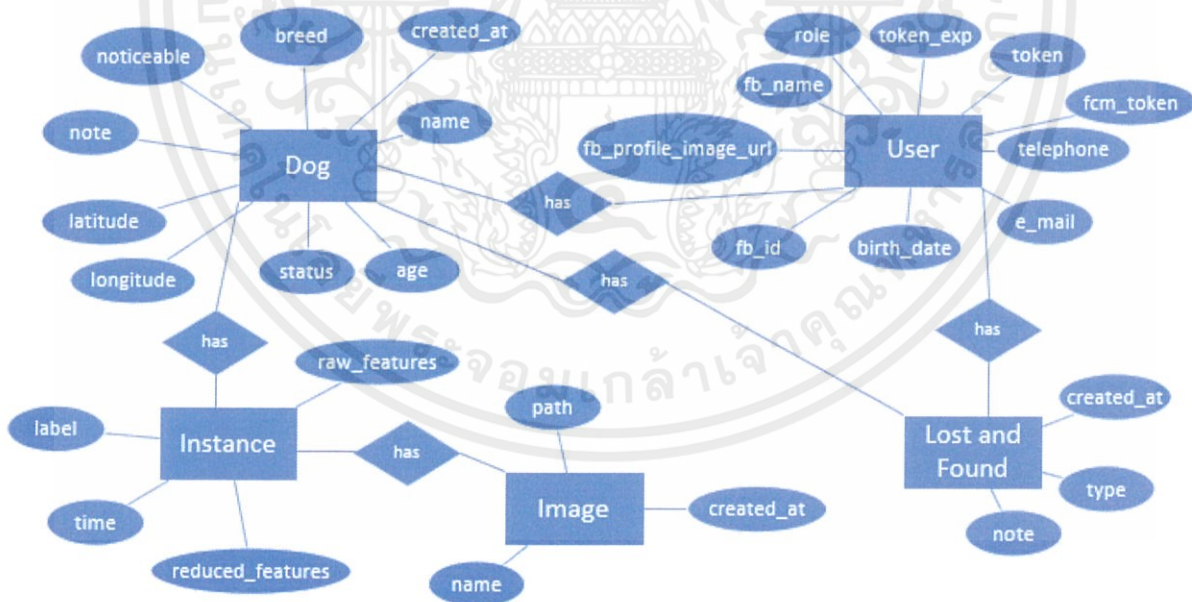
รูป 3.4 แผนผังการทำงานของอัลกอริทึมสำหรับค้นหารูปภาพใบหน้าสุนัขเหมือน

การทำงานของอัลกอริทึมแบ่งออกเป็น 6 ส่วน

- 1) ตัวดึงคุณสมบัติจากรูปภาพ (Feature Extractor) ทำหน้าที่ดึงคุณสมบัติออกมาจากรูปภาพที่ใส่เข้าไปซึ่งก็คือรูปภาพใบหน้าสุนัข โดยจะให้คุณสมบัติทั้งหมด 4,096 คุณสมบัติ
- 2) ตัวคัดเลือกคุณสมบัติ (Feature Selector) ทำหน้าที่กรองคุณสมบัติที่ไม่สำคัญหรือไม่มีน้ำหนักเพียงพอตามค่าที่กำหนด โดยข้อมูลชุดหนึ่งจะส่งให้ตัวคัดแยก (Classifier) อีกชุดหนึ่งจะส่งให้เครื่องมือค้นหารูปภาพสุนัขเหมือน (Nearest Neighbor Search Engine) ซึ่งข้อมูลทั้งสองชุดเป็นข้อมูลเดียวกัน
- 3) ตัวคัดแยก (Classifier) ทำหน้าที่คัดแยกคุณสมบัติที่ได้รับจากตัวคัดเลือกคุณสมบัติว่าอยู่ในกลุ่มใด โดยจะให้ชื่อกลุ่ม (class หรือ label) เป็นผลลัพธ์ออกมา

- 4) ตัวดึงคุณสมบัติของรูปภาพใบหน้าสุนัข (Dog Face Features Inquirer) ทำหน้าที่ค้นหา กลุ่มของคุณสมบัติที่ถูกระบุจากตัวคัดแยกในฐานข้อมูลใบหน้าสุนัข (Dog Faces Database) ซึ่งก็คือกลุ่มของใบหน้าสุนัขที่มีลักษณะคล้ายกัน โดยจะส่งให้เครื่องมือ ค้นหารูปภาพใบหน้าสุนัขเหมือน (Nearest Neighbor Search Engine)
- 5) เครื่องมือค้นหารูปภาพใบหน้าสุนัขเหมือน (Nearest Neighbor Search Engine) เป็น เครื่องมือเปรียบเทียบคุณสมบัติ โดยจะหาคุณสมบัติที่ได้จากตัวดึงคุณสมบัติของ รูปภาพใบหน้าสุนัขที่มีความใกล้เคียงกันกับคุณสมบัติที่ได้จากตัวคัดเลือกคุณสมบัติ มากที่สุดตามจำนวนที่ถูกกำหนด จากนั้นจะส่งรายชื่อของรูปภาพที่มีคุณสมบัติ ใกล้เคียงไปให้ตัวดึงรูปภาพใบหน้าสุนัข (Dog Face Photo Inquirer)
- 6) ตัวดึงรูปภาพใบหน้าสุนัข (Dog Face Photo Inquirer) จะทำหน้าที่ค้นหารูปภาพ ใบหน้าสุนัขจากฐานข้อมูลใบหน้าสุนัขตามที่กำหนดในรายชื่อที่ได้รับจากเครื่องมือ ค้นหารูปภาพใบหน้าสุนัขเหมือน และส่งกลุ่มของรูปใบหน้าสุนัขเหมือนออกเป็น ผลลัพธ์

3.5 ER diagram ของระบบ



รูป 3.5 แสดง ER diagram ของระบบ

ระบบแบ่งออกเป็น 5 entity ได้แก่

- 1) User Entity เป็น entity ที่แสดงข้อมูลที่จำเป็นทั้งหมดของผู้ใช้ เช่น ระดับของผู้ใช้ (role) โทเคนสำหรับติดต่อเครื่องมือสื่อสารของผู้ใช้ (fcm_token) นอกจากนี้ยังมีข้อมูลบางส่วนซึ่งมีการเชื่อมโยงกับข้อมูลของผู้ใช้บนสื่อสังคมออนไลน์ Facebook เช่น یدیชื่อ หรืออีเมลที่แสดงบน Facebook
- 2) Dog Entity เป็น entity ที่แสดงข้อมูลของสุนัข 1 ตัว เช่น ชื่อ จุดสังเกตบนลำตัว (noticeable) สายพันธุ์ (breed) ละติจูด ลองจิจูด อายุ และสถานะของสุนัขตัวดังกล่าว เช่น ยังอยู่กับเจ้าของ หรือสูญหาย
- 3) Instance Entity เป็น entity ที่แสดงข้อมูลของรูปภาพใบหน้าสุนัข โดยจะเก็บคุณสมบัติปกติ (raw features) คุณสมบัติที่ผ่านกระบวนการลดคุณสมบัติแล้ว (reduced features) กลุ่มของสุนัขที่มีลักษณะใกล้เคียง (label)
- 4) Image Entity เป็น entity ที่แสดงข้อมูลของรูปภาพ เช่น ชื่อ เวลาที่สร้าง และไคเรคทอรีที่บ่งบอกถึงที่อยู่ของภาพ
- 5) Lost and Found Entity เป็น entity ที่แสดงข้อมูลของประกาศพบสุนัขและประกาศสุนัขหาย ซึ่งจะถูกระบุด้วยชนิดของการประกาศ (type) และเก็บคำประกาศ (note)

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้เนื้อหาจะประกอบไปด้วยการทดลองและผลการทดลองในการทดสอบอัลกอริทึมของระบบและการทดลองใช้แอปพลิเคชัน โดยส่วนของการทดสอบอัลกอริทึมนั้นจะทำการเก็บรวบรวมข้อมูล การสร้างตัวคัดแยกไปจนถึงผลการทดลอง ส่วนของการทดลองใช้แอปพลิเคชัน จะแสดงขั้นตอนการทดลองและผลลัพธ์ที่เกิดขึ้น

4.1 การทดลองอัลกอริทึมของระบบ

4.1.1 การลดคุณสมบัติ (Feature Reduction)

4.1.1.1 การทดลองการลดคุณสมบัติด้วยการวิเคราะห์การถดถอยเชิงเส้น

4.1.1.1.1 วัตถุประสงค์

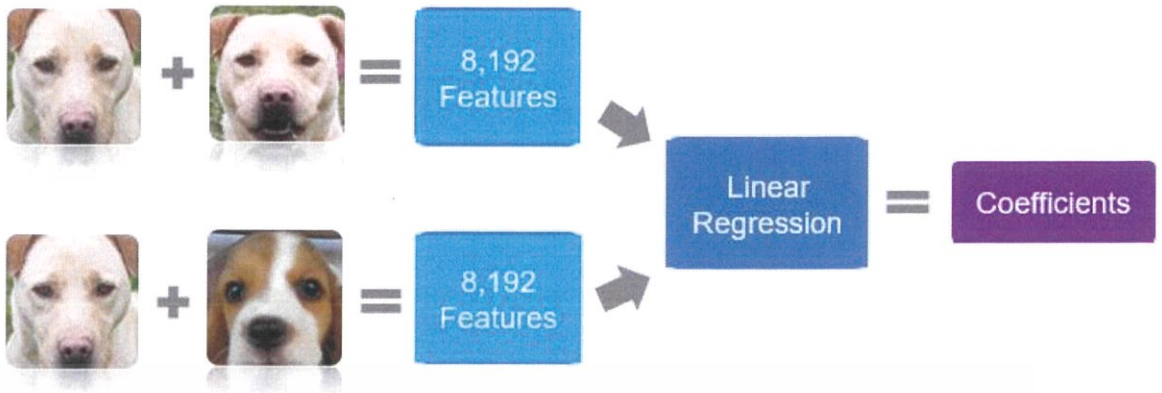
เพื่อลดจำนวนคุณสมบัติที่ไม่จำเป็นที่อาจก่อให้เกิดปัญหา Curse of Dimensionality ส่งผลให้ประสิทธิภาพโดยรวมการทำงานของอัลกอริทึมลดลง

4.1.1.1.2 สมมติฐาน

สามารถลดจำนวนคุณสมบัติลงได้อย่างน้อยร้อยละ 30 ถึงร้อยละ 50 จากคุณสมบัติจำนวนทั้งหมด 4,096 คุณสมบัติ โดยที่ค่าความแม่นยำ (accuracy) และค่าความบริสุทธิ์ (purity) มีค่าเท่าเดิม เพิ่มขึ้นหรือลดลงเล็กน้อย

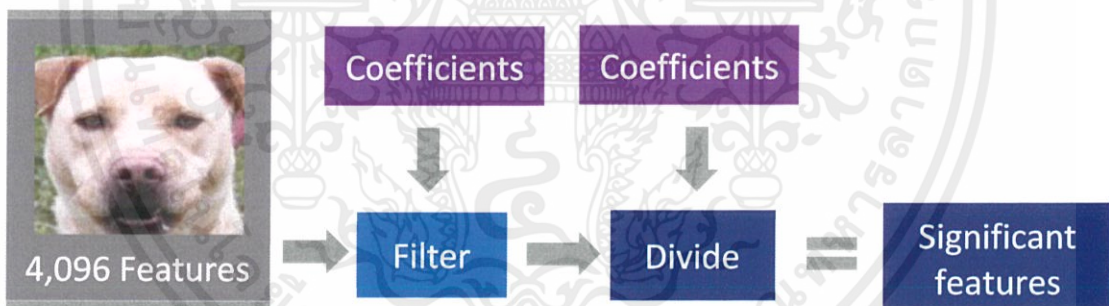
4.1.1.1.3 ขั้นตอนการทดลอง

ขั้นตอนแรกคือการดึงคุณสมบัติจากรูปใบหน้าสุนัขด้วย Overfeat ซึ่งจะได้คุณสมบัติ 4,096 จำนวน ขั้นตอนถัดไปคือการนำคุณสมบัติทั้งหมดที่ไม่ทราบความหมายมาลดจำนวนให้เหลือเฉพาะคุณสมบัติที่จำเป็น โดยใช้การวิเคราะห์การถดถอยเชิงเส้นมาช่วย โดยจับคู่คุณสมบัติภาพใบหน้าสุนัขพันธุ์เดียวกันและให้คลาสคำตอบเป็น 1 และจับคู่คุณสมบัติภาพใบหน้าสุนัขพันธุ์และให้คลาสคำตอบเป็น 0 จากนั้นจึงสุ่มการจับคู่โดยให้มีอัตราส่วนทั้ง 2 แบบเท่ากันแล้วจึงนำมาสร้าง model ซึ่งผลที่ได้คือสัมประสิทธิ์ (coefficient) ของแต่ละคุณสมบัติดังรูป 4.1



รูป 4.1 กระบวนการสร้างโมเดลด้วยอัลกอริทึม Linear Regression Analysis

จากนั้นนำคุณสมบัติมาผ่านการคัดกรองด้วยสัมประสิทธิ์ที่ได้จากขั้นตอนก่อนหน้า โดยคุณสมบัติใดที่สัมประสิทธิ์มีน้ำหนักต่อคลาสคำตอบต่ำจะถูกคัดออก ทิ้งเหลือไว้เฉพาะคุณสมบัติที่สัมประสิทธิ์มีน้ำหนักต่อคลาสคำตอบสูง ซึ่งคุณสมบัติที่เหลือจะถูกหารด้วยค่าสัมประสิทธิ์ของตัวเองเพื่อลดการกระจายตัวของค่าคุณสมบัติในปริภูมิ ดังรูป 4.2



รูป 4.2 แสดงการคัดกรองคุณสมบัติและการลดการกระจายค่าของคุณสมบัติในปริภูมิ

4.1.1.1.4 ข้อมูลที่ใช้ในการทดลอง

คุณสมบัติของรูปใบหน้าสุนัขจำนวน 30 สายพันธุ์ พันธุ์ละ 30 รูป รวม 900 รูป

4.1.1.1.5 ตัววัดผล

ในการทดลอง จะใช้ค่าความแม่นยำ ค่าความบริสุทธิ์และจำนวนคุณสมบัติหลังผ่านการทดลองเป็นตัววัดผล โดยค่าความแม่นยำและค่าความบริสุทธิ์จะต้องใกล้เคียงค่าก่อนการทดลองหรือมีค่าเพิ่มสูงขึ้น และจำนวนคุณสมบัติจะต้องลดลงจากเดิมอย่างน้อยร้อยละ

ละ 30

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1.1.6 ผลการทดลอง

จากผลการทดลอง สามารถคัดเลือคุณสมบัติเหลือเพียง 289 คุณสมบัติ จากเดิม 4,096 คุณสมบัติ หรือลดจำนวนคุณสมบัติไปได้ร้อยละ 92.94 ในส่วนของความแม่นยำมีค่าเพิ่มขึ้นจากเดิมร้อยละ 49.80 เป็นร้อยละ 67.25 ส่วนค่าความบริสุทธิ์ลดลงจากร้อยละ 77.68 เหลือร้อยละ 71.45

4.1.1.2 การทดลองการลดคุณสมบัติด้วยการวิเคราะห์องค์ประกอบหลัก

4.1.1.2.1 วัตถุประสงค์

เพื่อลดจำนวนคุณสมบัติให้เหลือเฉพาะคุณสมบัติที่จำเป็นเท่านั้น

4.1.1.2.2 สมมติฐาน

สามารถลดจำนวนคุณสมบัติลงได้อย่างน้อยร้อยละ 30 ถึงร้อยละ 50 จากคุณสมบัติจำนวนทั้งหมด 289 คุณสมบัติ ที่ได้จากการทดลองที่ 4.1.1.1 โดยที่ค่าความแม่นยำ (accuracy) และค่าความบริสุทธิ์ (purity) มีค่าเท่าเดิม เพิ่มขึ้นหรือลดลงเล็กน้อย

4.1.1.2.3 ขั้นตอนการทดลอง

นำคุณสมบัติของใบหน้าสุนัขทุกตัวมาสร้าง โมเดล โดยใช้อัลกอริทึมวิเคราะห์องค์ประกอบหลัก กำหนดให้อัตราการรักษาองค์ประกอบ (n -components) อยู่ที่ 0.9 จากนั้นนำโมเดลที่ได้มาแปลง (transform) ปริภูมิของแต่ละคุณสมบัติ ซึ่งปริภูมิใดมีขนาด eigen vector น้อยจะถูกโปรเจกตลงไปยังปริภูมิที่มีขนาด eigen vector สูง ทำให้จำนวนปริภูมิลดลง แต่คุณสมบัติของปริภูมิที่มีขนาด eigen vector สูงจะเปลี่ยนไปด้วยเช่นกัน ส่งผลให้จำนวนคุณสมบัติลดลงตามปริภูมิที่ถูกโปรเจกตลงไปยังปริภูมิอื่น

4.1.1.2.4 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลคุณสมบัติของใบหน้าสุนัขในระบบทั้งหมดจากการทดลอง 4.1.1.1

4.1.1.2.5 ตัววัดผล

ในการทดลอง จะใช้ค่าความแม่นยำ ค่าความบริสุทธิ์และจำนวนคุณสมบัติหลังผ่านการทดลองเป็นตัววัดผล โดยค่าความแม่นยำและค่าความบริสุทธิ์จะต้องใกล้เคียงค่าก่อนการทดลองหรือมีค่าเพิ่มสูงขึ้น และจำนวนคุณสมบัติจะต้องลดลงจากเดิมอย่างน้อยร้อยละ 30

4.1.1.2.6 ผลการทดลอง

จากผลการทดลอง สามารถลดคุณสมบัติเหลือเพียง 35 คุณสมบัติ จากเดิม 289 คุณสมบัติ ในการทดลองที่ 4.1.1.1 หรือลดจำนวนคุณสมบัติไปได้ร้อยละ 87.88 ใน

ส่วนของความแม่นยำและค่าความบริสุทธิ์มีค่าเท่าเดิมคือร้อยละ 67.25 และร้อยละ 71.45 ตามลำดับ

4.1.2 การจัดกลุ่ม (Clustering)

การทดลองชุดนี้มีวัตถุประสงค์เพื่อค้นหาอัลกอริทึมที่เหมาะสมกับกระบวนการจัดกลุ่มใบหน้าสุนัข โดยจะเปรียบเทียบระหว่างอัลกอริทึม K-Means ซึ่งเป็นอัลกอริทึมที่ต้องมีการกำหนดจำนวนกลุ่มหรือค่า k กับอัลกอริทึม Affinity Propagation ซึ่งเป็นอัลกอริทึมที่กำหนดจำนวนกลุ่มให้โดยอัตโนมัติ โดยแบ่งออกเป็น 2 การทดลอง ได้แก่

4.1.2.1 การทดลองหาจำนวนกลุ่มที่เหมาะสมที่สุดของอัลกอริทึม K-Means ในการจัดกลุ่มใบหน้าสุนัข

4.1.2.1.1 วัตถุประสงค์

เพื่อค้นหาจำนวนกลุ่มหรือค่า k ที่เหมาะสมที่สุดของอัลกอริทึม K-Means ในการจัดกลุ่มใบหน้าสุนัขที่ให้ค่าความแม่นยำและค่าความบริสุทธิ์สูงสุด

4.1.2.1.2 สมมติฐาน

อัลกอริทึมนี้สามารถจัดกลุ่มใบหน้าสุนัขได้อย่างเหมาะสม มีความแม่นยำและความบริสุทธิ์อยู่ในช่วงหรือเกินร้อยละ 75 ถึงร้อยละ 85

4.1.2.1.3 ขั้นตอนการทดลอง

ขั้นตอนแรกคือการแบ่งชุดข้อมูลใบหน้าสุนัขทั้งหมดออกเป็น 2 ชุด ได้แก่ชุดข้อมูลทดสอบและชุดข้อมูลฝึก โดยนำชุดข้อมูลฝึกมาสร้างโมเดลโดยใช้อัลกอริทึม K-Means ซึ่งการใช้งานอัลกอริทึมดังกล่าวจะต้องมีการกำหนดค่า k เสมอ เพื่อบอกจำนวนกลุ่มที่ต้องการ แต่การกำหนดค่า k สามารถทำได้หลากหลาย จึงต้องมีการเลือกค่า k ที่ให้ผลการทดลองที่ดีที่สุด โดยการเพิ่มจำนวนค่า k ทีละ 1 จนถึง 15 จากนั้นจึงนำชุดข้อมูลทดสอบมาผ่านโมเดลและสังเกตผลค่าความแม่นยำและค่าความบริสุทธิ์

4.1.2.1.4 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลคุณสมบัติของใบหน้าสุนัขในระบบทั้งหมดที่ผ่านการทดลอง 4.1.1.2

4.1.2.1.5 ตัววัดผล

ในการทดลองจะใช้ค่าความแม่นยำ ค่าความบริสุทธิ์ และค่า F_{Measure} มาเป็นตัววัด โดยค่าความแม่นยำและค่าความบริสุทธิ์จะต้องสูงที่สุดในจำนวนกลุ่มที่ถูกกำหนด แต่เนื่องจากค่าความแม่นยำมีแนวโน้มสวนทางกับค่าความบริสุทธิ์ ดังนั้นค่า F_{Measure} จึงสามารถนำมาช่วยตัดสินใจจำนวนกลุ่มที่เหมาะสมที่ให้ค่าความแม่นยำและค่าความ

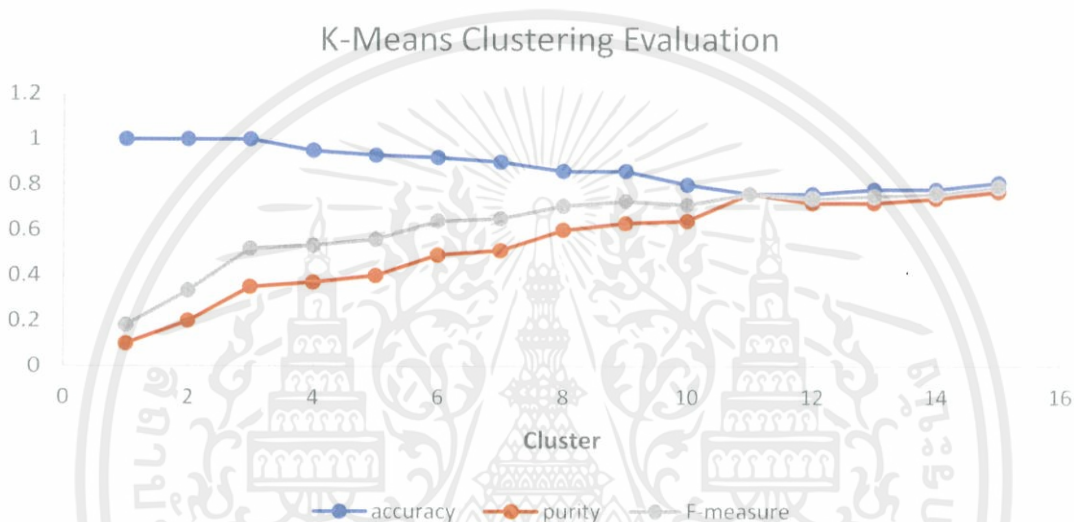
บริสุทธิ์สูงทั้งสองค่าได้ โดยนำค่าทั้งสองมาคำนวณ ยิ่งค่า $F_{Measure}$ สูง ค่าความแม่นยำ และค่าความบริสุทธิ์จะยิ่งสูงตาม โดยสูตรการคำนวณค่า $F_{Measure}$ คือ

$$F_{Measure} = \frac{2PR}{P + R}$$

เมื่อ P คือ Precision หรือค่าความระลึกลับ ซึ่งในที่นี้คือค่าความบริสุทธิ์ (purity)

R คือ Recall หรือค่าความแม่นยำ ซึ่งในที่นี้คือค่าความแม่นยำที่วัดได้

4.1.2.1.4 ผลการทดลอง



รูป 4.3 กราฟแสดงผลการทดลองการจัดกลุ่มด้วยอัลกอริทึม K-Means

จากผลการทดลองรูปที่ 4.3 กราฟแสดงค่าความแม่นยำมีแนวโน้มลดลงเมื่อจำนวนกลุ่ม (cluster) หรือค่า k เพิ่มขึ้น สาเหตุเกิดจากการที่มีจำนวนกลุ่มน้อย โอกาสที่อัลกอริทึมจะคัดแยกได้ถูกต้องมีโอกาสูงขึ้น หากมี 2 กลุ่ม ค่าความแม่นยำขั้นต่ำจะเท่ากับร้อยละ 50 เป็นต้น ในส่วนของค่าความบริสุทธิ์มีแนวโน้มสูงขึ้นเมื่อมีจำนวนกลุ่มมากขึ้น สาเหตุเกิดจากจำนวนกลุ่มแปรผันตรงกับค่าความบริสุทธิ์ตามสูตรการคำนวณหาค่าความบริสุทธิ์ ยิ่งมีจำนวนกลุ่มมาก ค่าความบริสุทธิ์ก็จะมากขึ้นตามไปด้วย ซึ่งไม่ใช่แนวทางปฏิบัติที่ดี เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด เราสามารถใช้ค่า F Measure ช่วยตัดสินใจได้ โดยเลือกจำนวนกลุ่มที่ให้ค่า $F_{Measure}$ มากที่สุดซึ่งก็คือ 15 กลุ่ม โดยให้ค่าความแม่นยำที่ร้อยละ 81 และให้ค่าความบริสุทธิ์ที่ร้อยละ 77

4.1.2.2 การทดลองการจัดกลุ่มด้วยอัลกอริทึม Affinity Propagation เทียบกับอัลกอริทึม

K-Means

4.1.2.2.1 วัตถุประสงค์

เพื่อทำการเปรียบเทียบประสิทธิภาพการจัดกลุ่มด้วยอัลกอริทึม Affinity Propagation ซึ่งมีการจัดกลุ่มให้อัตโนมัติกับอัลกอริทึม K-Means ซึ่งต้องมีการกำหนดจำนวนกลุ่มก่อน

4.1.2.2.2 สมมติฐาน

อัลกอริทึม Affinity Propagation สามารถแบ่งกลุ่มให้อัตโนมัติ โดยไม่ต้องกำหนดจำนวนกลุ่มล่วงหน้า มีค่าความแม่นยำและค่าความบริสุทธิ์ใกล้เคียงกับผลจากอัลกอริทึม K-Means หรือให้ประสิทธิภาพที่ดีกว่า

4.1.2.2.3 ขั้นตอนการทดลอง

ขั้นตอนแรกคือการแบ่งชุดข้อมูลใบหน้าสุนัขทั้งหมดออกเป็น 2 ชุด ได้แก่ชุดข้อมูลทดสอบและชุดข้อมูลฝึก โดยนำชุดข้อมูลฝึกมาสร้างโมเดลโดยใช้อัลกอริทึม Affinity Propagation จากนั้นนำชุดข้อมูลทดสอบมาผ่านอัลกอริทึมเพื่อวัดผล เทียบกับอัลกอริทึม K-Means

4.1.2.2.4 ข้อมูลที่ใช้ในการทดลอง

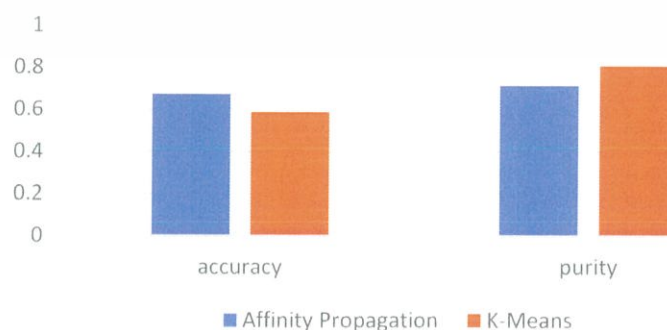
ข้อมูลคุณสมบัติของใบหน้าสุนัขในระบบทั้งหมดที่ผ่านการทดลอง 4.1.1.2 ค่าความแม่นยำและค่าความบริสุทธิ์ของอัลกอริทึม K-Means ที่ได้จากการทดลอง 4.1.2.1

4.1.2.2.5 ตัววัด

ค่าความแม่นยำและค่าความบริสุทธิ์ที่ได้จากอัลกอริทึม Affinity Propagation และอัลกอริทึม K-Means

4.1.2.2.6 ผลการทดลอง

Efficiency Comparison of Algorithms



รูป 4.4 แสดงผลการทดลองเปรียบเทียบ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระหว่างอัลกอริทึม Affinity Propagation และ K-Means ผลจากการทดลอง อัลกอริทึม Affinity Propagation แบ่งกลุ่มให้ 11 กลุ่ม โดยอัตโนมัติ และจากรูปที่ 4.4 พบว่ามีค่าความแม่นยำมากกว่าอัลกอริทึม K-Means ที่ร้อยละ 67.25 แต่มีค่าความบริสุทธิ์น้อยกว่าที่ร้อยละ 71.45 ซึ่งโดยภาพรวม อัลกอริทึม Affinity Propagation มีผลลัพธ์ที่ดีกว่า เนื่องจากไม่ต้องกำหนดจำนวนกลุ่มล่วงหน้า มีค่าความแม่นยำมากกว่า และมีค่าความบริสุทธิ์ต่างกันเล็กน้อย

สุนัขบางตัวไม่สามารถระบุชื่อพันธุ์ได้ เนื่องจากมีลักษณะผสมร่วมหลายพันธุ์ บางครั้งไม่สามารถระบุได้ว่าเป็นสายพันธุ์ใด เมื่อผ่านการจัดกลุ่มใบหน้าสุนัขด้วย อัลกอริทึม Affinity Propagation จะได้สุนัขที่แบ่งตามกลุ่มดังรูป 4.5 ซึ่งเป็นตัวอย่างเพียงบางส่วนของรูปใบหน้าสุนัขทั้งระบบ จะสามารถสังเกตได้ว่า ภาพใบหน้าของสุนัขแต่ละกลุ่ม ไม่ได้มีการคัดแยกตามสายพันธุ์



กลุ่มที่ 1



กลุ่มที่ 2



กลุ่มที่ 3



กลุ่มที่ 4

รูป 4.5 ผลลัพธ์เมื่อผ่าน Clustering

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 อัลกอริทึมค้นหาใบหน้าสุนัข

การทดลองชุดนี้มีจุดประสงค์เพื่อทดสอบการทำงานของอัลกอริทึมค้นหาใบหน้าสุนัขซึ่งจะค้นหาจากใบหน้าสุนัขในระบบทั้งหมด และให้ผลลัพธ์ออกมาเป็นใบหน้าสุนัขที่มีความใกล้เคียง 5 อันดับแรก

4.1.3.1 การทดสอบอัลกอริทึม Nearest Neighbor Search

4.1.3.1.1 วัตถุประสงค์

เพื่อให้สามารถใส่รูปภาพอินพุตลงไปแล้วอัลกอริทึมสามารถค้นหารูปใบหน้าสุนัขที่คล้ายกับรูปภาพอินพุตได้

4.1.3.1.2 สมมติฐาน

อัลกอริทึมสามารถค้นหาใบหน้าสุนัขที่มีความใกล้เคียงได้อย่างน้อยเป็นสายพันธุ์เดียวกันหรือมีรูปร่างลักษณะ ลวดลายหรือสีที่คล้ายกัน

4.1.3.1.3 ขั้นตอนการทดลอง

นำรูปภาพอินพุตมาผ่านกระบวนการดึงคุณสมบัติ กระบวนการลดคุณสมบัติ และกระบวนการคัดแยก เพื่อค้นหาว่ามีลักษณะใกล้เคียงกับกลุ่มใด จากนั้นนำคุณสมบัติภาพใบหน้าของสุนัขในกลุ่มนั้นทั้งหมดรวมทั้งคุณสมบัติของรูปภาพอินพุตมาสร้างปริภูมิ หลังจากนั้นทำการหา Euclidean distance โดยให้รูปภาพอินพุตเป็นจุดศูนย์กลาง กำหนดให้รัศมียูคลิดีเดนไม่เกิน 10 และผลลัพธ์ไม่เกิน 5 ผลลัพธ์ที่ได้จะเป็นใบหน้าสุนัขที่คุณสมบัติมีลักษณะคล้ายกับคุณสมบัติของรูปอินพุตมากที่สุดเรียงลำดับหนึ่งถึงห้า

4.1.3.1.4 ข้อมูลที่ใช้ในการทดลอง

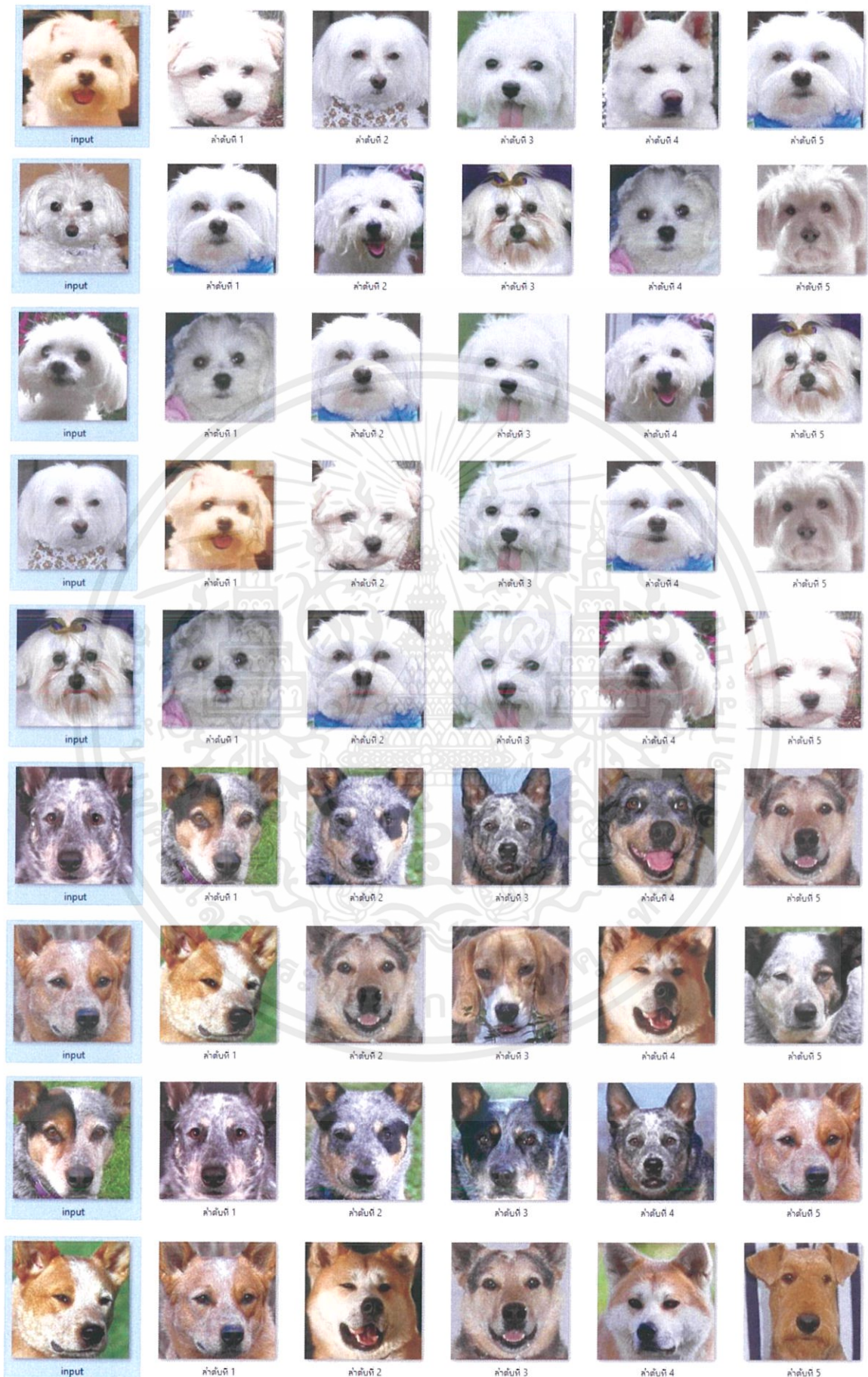
รูปภาพอินพุต คุณสมบัติของรูปภาพอินพุต รูปภาพใบหน้าสุนัขและคุณสมบัติของรูปภาพใบหน้าสุนัขทั้งหมดในระบบ โดยรูปภาพอินพุตจะทำการสุ่มดึงออกมาจากระบบจำนวน 6 พันธุ์ พันธุ์ละ 2 – 5 รูป รวมเป็น 21 รูป หลังการทดสอบเสร็จในแต่ละครั้ง จะใส่รูปภาพอินพุตกลับเข้าไปในระบบ

4.1.3.1.5 ตัววัด

อันดับความคล้ายของรูปภาพผลลัพธ์ต่อรูปภาพอินพุต

4.1.3.1.6 ผลการทดลอง

จากการทดลองนำสุนัขจำนวน 21 ตัว (input) ไปทดลองในระบบให้ผลลัพธ์ของการเรียงลำดับจาก ลำดับที่ 1 ถึง ลำดับที่ 5 โดยเรียงตามความคล้ายจากซ้ายไปขวาตามลำดับดังรูป 4.6



เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4.6 ผลลัพธ์สุดท้ายของการทดลอง

4.1.3.2 การทดสอบความพึงพอใจการทำงานของอัลกอริทึมค้นหาใบหน้าสุนัขเหมือน โดยผู้ใช้

4.1.3.2.1 วัตถุประสงค์

เพื่อเป็นการช่วยประเมินการทำงานของอัลกอริทึมในมุมมองของผู้ใช้ นอกเหนือจากการประเมินด้วยค่าความแม่นยำและค่าความบริสุทธิ์ ทำให้ อัลกอริทึมถูกประเมินการทำงานอย่างรอบด้าน

4.1.3.2.2 สมมติฐาน

ได้คะแนนความพึงพอใจขั้นต่ำเฉลี่ย 3.5 คะแนน เต็ม 5 คะแนน

4.1.3.2.3 ขั้นตอนการทดลอง

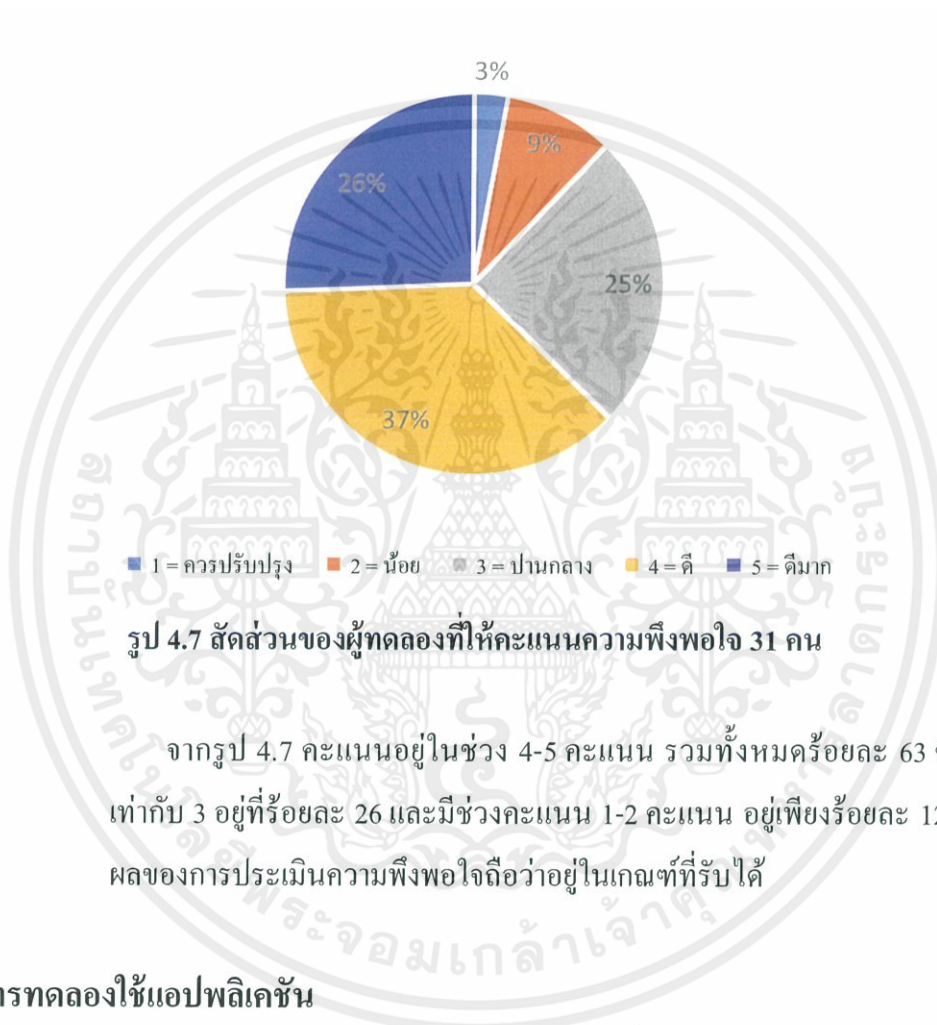
ทำการสุ่มรูปภาพอินพุตตัวอย่างจากระบบ 21 รูปภาพ จากนั้นนำมาผ่าน อัลกอริทึมซึ่งจะให้ผลลัพธ์ออกมาเป็นใบหน้าสุนัขคล้ายจำนวน 5 ใบหน้า ของแต่ละรูปภาพอินพุต นำรูปภาพทั้งหมดมาสร้างเป็นแบบสอบถามจำนวน 21 คำถาม ตามจำนวนรูปภาพอินพุต แสดงให้ผู้ใช้งานเห็นถึงการเรียงลำดับความเหมือนของ รูปภาพ ผู้ใช้จะต้องให้คะแนนความพึงพอใจในแต่ละกลุ่ม โดยแบ่งคะแนนเป็น ช่วงตามความพึงพอใจได้แก่ 5 หมายถึง พึงพอใจมากที่สุด 4 หมายถึง พึงพอใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มาก 3 หมายถึง พึงพอใจปานกลาง 2 หมายถึง พึงพอใจน้อย 1 หมายถึง ควรปรับปรุง หลังจากนั้นทำการรวบรวมคะแนนจากแบบสอบถามทั้งหมดและหาคะแนนเฉลี่ยรวม

4.1.3.2.4 ผลการทดลอง

เมื่อหาค่าเฉลี่ยของคะแนนแล้ว ผลที่ได้คือ 3.73 ซึ่งอยู่ในช่วงความพึงพอใจปานกลางไปถึงความพึงพอใจมาก



จากรูป 4.7 คะแนนอยู่ในช่วง 4-5 คะแนน รวมทั้งหมดร้อยละ 63 คะแนนเท่ากับ 3 อยู่ที่ร้อยละ 26 และมีช่วงคะแนน 1-2 คะแนน อยู่เพียงร้อยละ 12 ดังนั้นผลของการประเมินความพึงพอใจถือว่าอยู่ในเกณฑ์ที่รับได้

4.2 การทดลองใช้แอปพลิเคชัน

หลังจากการพัฒนาระบบตามการออกแบบในบทที่ 3 แล้วขั้นตอนถัดมาคือการทดสอบระบบ วัตถุประสงค์คือ ทดสอบการทำงานของแอปพลิเคชันและเซิร์ฟเวอร์ให้เป็นไปอย่างถูกต้องตามที่ต้องการ

4.2.1. การเข้าสู่ระบบ

4.2.1.1 วัตถุประสงค์

เพื่อทดสอบการเข้าสู่ระบบของผู้ใช้และการร้องขอข้อมูลจาก Facebook และการส่งข้อมูลไปเก็บที่เซิร์ฟเวอร์ของแอปพลิเคชัน

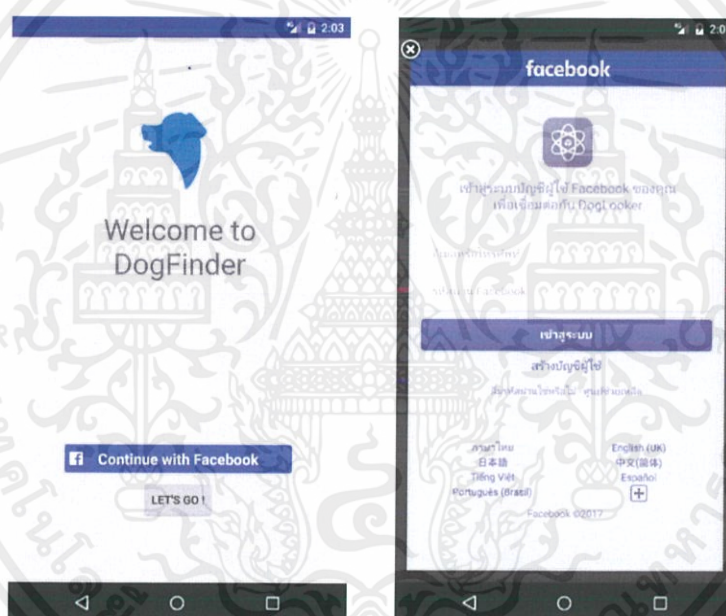
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1.2 ขั้นตอนการทดลอง

ติดตั้งแอปพลิเคชันลงอุปกรณ์และเปิดใช้งานแอปพลิเคชัน ก่อนที่ผู้ใช้จะใช้งานแอปพลิเคชันจะต้องทำการล็อกอินระบบก่อนทุกครั้ง โดยคลิกที่ปุ่ม Continue with Facebook ดังรูป 4.8 (ก) หากผู้ใช้ยังไม่ได้เป็นสมาชิกของ Facebook ก็สามารถกดที่ปุ่มนี้ได้เช่นกัน หากผู้ใช้เป็นสมาชิกของ Facebook อยู่แล้วให้กรอก Email และ Password ของ Facebook หากผู้ใช้ไม่ได้เป็นสมาชิกของ Facebook ให้กดที่ สร้างบัญชีผู้ใช้จากนั้นกรอกข้อมูลส่วนตัวดังรูป 4.8 (ข)

4.2.1.3 ผลการทดลอง

ระบบสามารถดึงข้อมูลจาก Facebook และจัดเก็บไว้บนเซิร์ฟเวอร์ได้อย่างถูกต้อง



(ก)

(ข)

รูป 4.8 ส่วนแสดงผลการเข้าสู่ระบบ

(ก) ส่วนแรกของแอปพลิเคชัน

(ข) ส่วนของการเข้าสู่ระบบโดยอีเมลและรหัสผ่าน

4.2.2. การเพิ่มและแสดงรายการสุนัขของผู้ใช้

4.2.2.1. วัตถุประสงค์

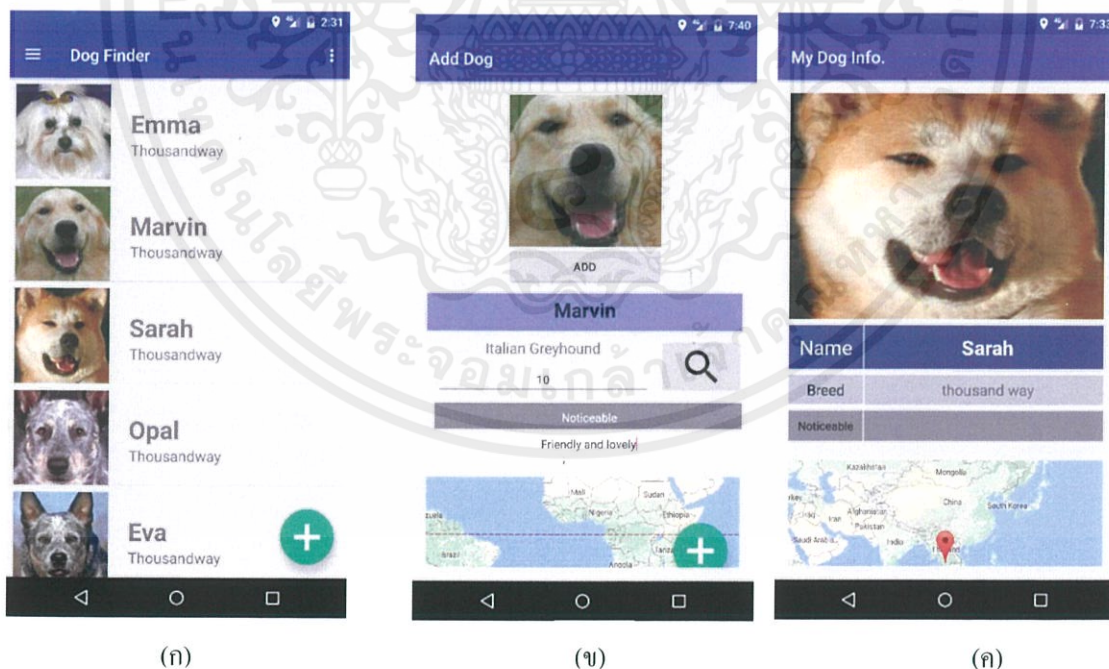
เพื่อทดสอบการแสดงผลรายการสุนัขที่ถูกเพิ่มขึ้นมา โดยการส่งการร้องขอรายการของประกาศไปที่เซิร์ฟเวอร์ แอปพลิเคชันจะต้องแสดงผลรายการสุนัขที่มีอยู่ในระบบและตรวจสอบข้อมูลที่ครบถ้วนหลังจากเพิ่มสุนัข

4.2.2.2. ขั้นตอนการทดลอง

ก่อนที่ผู้ใช้จะใช้งานแอปพลิเคชัน ปุ่มสั่งการบนหน้านี้ประกอบไปด้วยปุ่มเพิ่มรูปเครื่องหมายบวกสีเขียวด้านขวาล่างเป็นปุ่มสำหรับเพิ่มสุนัขสำหรับเจ้าของที่ได้รับเลี้ยงหรือดูแลอยู่ ดังรูป 4.9 (ก) จากนั้นทำการทดสอบโดยการใส่ชื่อ สายพันธุ์ ลักษณะเด่นหรือจุดสังเกตและตำแหน่งบนแผนที่ให้ครบถ้วนดังรูป 4.9 (ข) นอกจากนี้เมื่อกดเข้าไปดูข้อมูลของสุนัขแต่ละตัว ผลการแสดงผลจะเป็นดังรูป 4.9 (ค) ซึ่งจะแสดงข้อมูลชื่อ สายพันธุ์ ลักษณะเด่นหรือจุดสังเกตและตำแหน่งบนแผนที่

4.2.2.3. ผลการทดลอง

แอปพลิเคชันแสดงผลรายการสุนัขได้อย่างถูกต้องและเมื่อทำการเพิ่มสุนัข สามารถแสดงข้อมูลได้อย่างครบถ้วน



รูป 4.9 ส่วนแสดงผลรายการสุนัขของผู้ใช้

(ก) ส่วนแสดงผลรายการสุนัขของผู้ใช้

(ข) ส่วนการเพิ่มสุนัข

(ค) ส่วนแสดงข้อมูลของสุนัข

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3. การเพิ่มและแสดงรายการประกาศสุนัขสูญหาย

4.2.3.1. วัตถุประสงค์

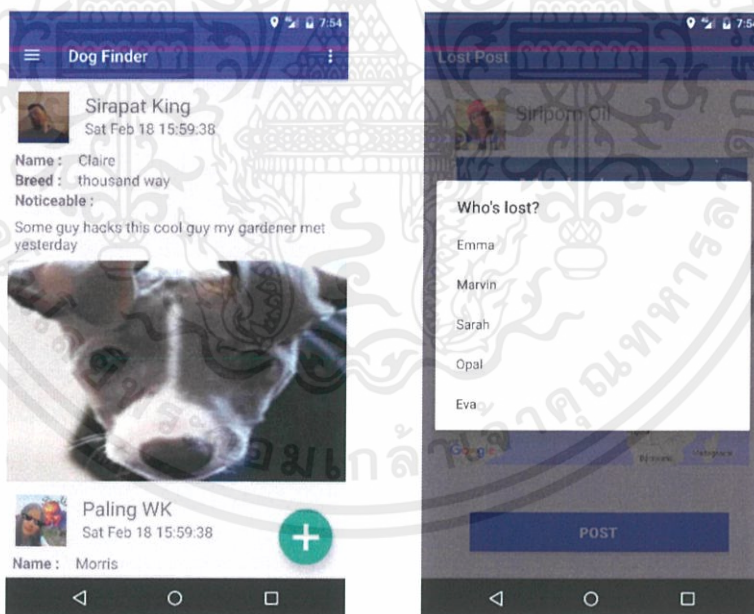
เพื่อทดสอบการเพิ่มข้อมูลและการแสดงรายการประกาศสุนัขหาย โดยการส่งการร้องขอรายการของประกาศไปที่เซิร์ฟเวอร์ แอปพลิเคชันจะต้องแสดงรายการประกาศสุนัขหายที่มีอยู่ในระบบให้แก่ผู้ใช้งาน

4.2.3.2. ขั้นตอนการทดลอง

การประกาศสุนัขสูญหายประกอบด้วย 2 ส่วน คือ ประกาศเพื่อค้นหาสุนัขของตนเองและเข้าไปดูประกาศเพื่อช่วยตามหาสุนัขที่หายไปของผู้ใช้ท่านอื่น โดยจะแสดงรายการของสุนัขทั้งหมดที่เจ้าของในระบบที่ทำกรประกาศไว้ดังรูป 4.10 (ก) ในส่วนของการเพิ่มประกาศ ระบบจะให้เลือกว่าสุนัขตัวใดของผู้ใช้หายไปดังรูป 4.10 (ข)

4.2.3.3. ผลการทดลอง

แอปพลิเคชันแสดงรายการประกาศสุนัขหายได้อย่างถูกต้องและเมื่อทำการเพิ่มประกาศ สามารถแสดงข้อมูลได้อย่างครบถ้วน



(ก)

(ข)

รูป 4.10 ส่วนแสดงประกาศสุนัขสูญหาย

(ก) ส่วนแสดงประกาศสุนัขสูญหาย

(ข) ส่วนการเพิ่มประกาศ

4.2.4. การเพิ่มและแสดงรายการประกาศพบเจอสุนัข

4.2.4.1. วัตถุประสงค์

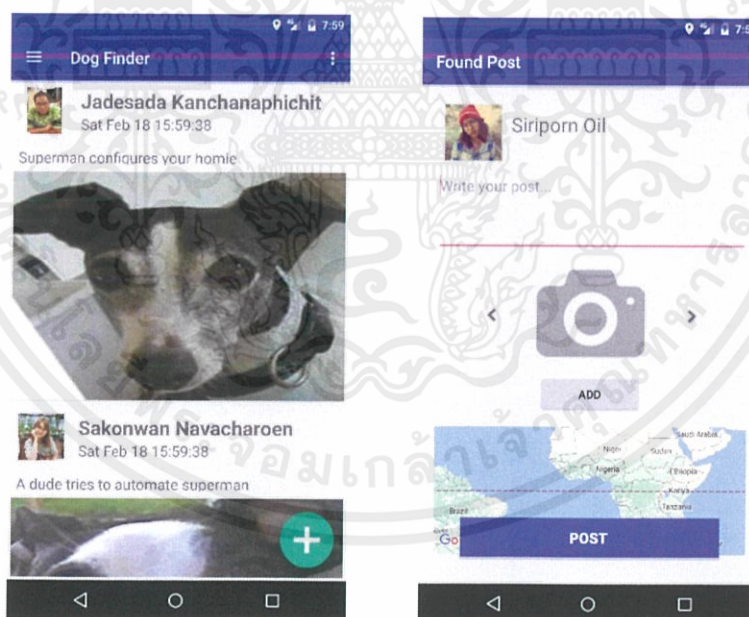
เพื่อทดสอบการเพิ่มข้อมูลและการแสดงรายการประกาศพบสุนัข โดยการส่งการร้องขอรายการของประกาศไปที่เซิร์ฟเวอร์ แอปพลิเคชันจะต้องแสดงรายการประกาศพบสุนัขที่มีอยู่ในระบบให้แก่ผู้ใช้งาน

4.2.4.2. ขั้นตอนการทดลอง

ทำการเพิ่มสุนัขไว้บ้านเหล่านี้เข้าไปในระบบ จากรูป 4.11 (ก) เมื่อกดเครื่องหมายบวกจะสามารถเพิ่มสุนัขที่พบเห็นได้ จากนั้นทำการกรอกข้อมูลในรูป 4.11 (ข) ให้ครบถ้วน เมื่อประกาศว่าสุนัขตัวใดตัวหนึ่งของผู้ใช้สูญหายแล้ว ผู้ใช้จะสามารถเข้าไปค้นหาสุนัขที่หน้าตาคล้ายคลึงกับสุนัขที่หายไปได้

4.2.4.2. ผลการทดลอง

แอปพลิเคชันแสดงรายการประกาศพบเจอสุนัขได้อย่างถูกต้องและเมื่อทำการเพิ่มสุนัข สามารถแสดงข้อมูลได้อย่างครบถ้วน



(ก)

(ข)

รูป 4.11 ส่วนแสดงประกาศพบเจอสุนัข

(ก) ส่วนแสดงประกาศพบเจอสุนัข

(ข) ส่วนการเพิ่มประกาศพบสุนัข

4.2.4. การประมวลผลใบหน้าของสุนัขที่คล้ายคลึงกัน

4.2.4.1. วัตถุประสงค์

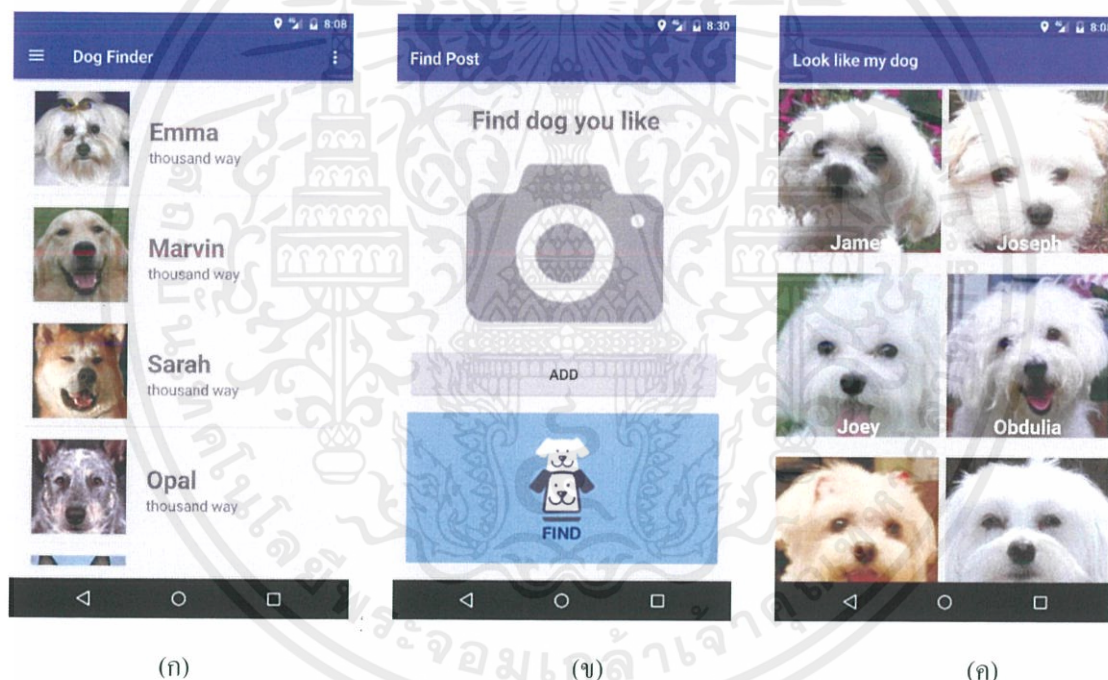
เพื่อทดสอบความถูกต้องของการแสดงรายการลำดับของรูปภาพที่ระบบประมวลผล

4.2.4.2. ขั้นตอนการทดลอง

ผู้ใช้งานสามารถค้นหาสุนัขที่มีใบหน้าคล้ายกับสุนัขของตนเองได้ ดังรูป 4.12 (ก) หรือจะเพิ่มรูปภาพสุนัขที่มีลักษณะที่ผู้ใช้ชื่นชอบเพื่อทำการรับเลี้ยงสุนัขดังกล่าว ดังรูป 4.12 (ข) ทั้งสองรูปแบบจะมีผลลัพธ์ที่แสดงสุนัขใบหน้าที่คล้ายคลึงกับรูปภาพที่ต้องการค้นหา ดังรูป 4.1 (ค)

4.2.4.3. ผลการทดลอง

แอปพลิเคชันแสดงรายการสุนัขได้อย่างถูกต้อง



รูป 4.12 ส่วนแสดงการค้นหาสุนัขใบหน้าคล้าย

- (ก) ส่วนแสดงรายการสุนัขของผู้ใช้ที่สูญหาย
- (ข) ส่วนการค้นหาสุนัขหน้าคล้ายที่ชื่นชอบ
- (ค) ส่วนแสดงผลลัพธ์ของการเลือกรายการที่ 1 ของ (ก)

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

DogFinder เป็นแอปพลิเคชันสำหรับค้นหาสุนัขสูญหายซึ่งประกอบด้วยหลายขั้นตอนของกระบวนการทำงาน ได้แก่ การดึงคุณสมบัติที่สำคัญจากรูปภาพเพื่อนำมาประมวลผล การคัดเลือกคุณสมบัติที่ไม่มีนัยสำคัญต่อการทำงานออกเพื่อลดเวลาในการประมวลผลและเพิ่มประสิทธิภาพการทำงาน การจัดกลุ่มใบหน้าสุนัขตามลักษณะความใกล้เคียงของใบหน้าโดยอาศัยคุณสมบัติที่ได้จากรูปภาพเพื่อให้สะดวกต่อการค้นหาเฉพาะกลุ่มแทนการค้นหาทั้งระบบ และการค้นหาเพื่อนบ้านใกล้เคียงหรือการค้นหาตามความคล้ายเพื่อให้ได้ใบหน้าที่มีความใกล้เคียงมากที่สุด ซึ่งนำมาประยุกต์ใช้งานเพื่อให้สามารถลดเวลาสำหรับใช้ในการค้นหาด้วยวิธีการแบบเดิม เช่น การประกาศตามที่สาธารณะ ซึ่งจะช่วยอำนวยความสะดวกให้แก่ผู้ใช้งานและลดจำนวนประชากรสุนัขไร้บ้าน

เนื่องจากแอปพลิเคชันประกอบไปด้วยการทำงานหลายกระบวนการดังที่กล่าวมา แต่ละกระบวนการจึงต้องมีการทดลองเพื่อคัดเลือกวิธีการและอัลกอริทึมที่มีการทำงานเหมาะสมที่สุด ได้แก่ การคัดเลือกอัลกอริทึมเพื่อลดจำนวนคุณสมบัติของรูปภาพซึ่งสามารถลดไปได้จากคุณสมบัติจำนวน 4096 คุณสมบัติ เหลือ 35 คุณสมบัติ หรือลดไปได้ร้อยละ 99.15 การคัดเลือกอัลกอริทึมเพื่อจัดกลุ่มและคัดแยกใบหน้าสุนัขซึ่งให้ค่าความแม่นยำสูงสุดร้อยละ 67.25 และการคัดเลือกอัลกอริทึมเพื่อค้นหาใบหน้าสุนัขที่มีความใกล้เคียงที่สุด ซึ่งระบบสามารถแสดงรูปภาพใบหน้าของสุนัขที่มีความคล้ายคลึงเรียงลำดับความเหมือนจากมากไปน้อยได้และสามารถทำงานได้เป็นอย่างดี แต่ยังคงมีความผิดพลาดในการทำงานอยู่ตามค่าความแม่นยำของอัลกอริทึมจัดกลุ่มและคัดเลือกใบหน้าสุนัขที่ร้อยละ 32.75 โดยจะสังเกตการเรียงลำดับได้ว่าบางครั้งอาจมีความคล้ายคลึงกันทั้งหมด แต่ในบางลำดับอาจมีความคล้ายคลึงมากกว่าแต่มีการเรียงลำดับที่ไม่ถูกต้องหรือมีภาพใบหน้าของสุนัขต่างสายพันธุ์ผสมเข้ามาในการเรียงลำดับ นอกจากนี้ยังมีการสำรวจความพึงพอใจผลลัพธ์ที่ได้จากการทำงานของระบบเพื่อให้ผู้ใช้เป็นคนประเมินว่าระบบมีประสิทธิภาพการทำงานที่เพียงพอหรือไม่ ซึ่งจากผลการประเมินพบว่าผู้ใช้ให้คะแนนเฉลี่ย 3.7 เต็ม 5 คะแนน แสดงให้เห็นถึงการทำงานของระบบมีประสิทธิภาพและได้รับความพึงพอใจจากผู้ใช้เป็นส่วนใหญ่

5.2 ปัญหาอุปสรรคและแนวทางแก้ไข

- 1) รูปภาพมีขนาดเล็กทำให้มีความ blur มากจนไม่สามารถปรับความละเอียดได้เพียงพอกับความต้องการ ซึ่งแนวทางแก้ไขคือต้องถ่ายรูปใบหน้าสุนัขใหม่เท่านั้น
- 2) กระบวนการดึงคุณสมบัติของรูปภาพในแต่ละรูปใช้พลังในการประมวลผลสูงและใช้เวลานาน ทำให้รองรับงานในเวลาเดียวกันได้น้อย ส่งผลให้ผู้ใช้ต้องรอเป็นเวลานานเมื่อมีการใช้งานพร้อมกันเป็นจำนวนมาก โดยแนวทางแก้ไขคือการทำ scale up ให้กับเซิร์ฟเวอร์ที่ให้บริการ และการปรับปรุงโปรแกรมให้มีการทำงานเป็นแบบคู่ขนานเพื่อรองรับการทำ scale out ซึ่งจะสามารถนำเซิร์ฟเวอร์หลายเครื่องมาช่วยในการประมวลผลและลดระยะเวลาการประมวลผลลงได้
- 3) อาจมีรูปภาพสุนัขที่หน้าไม่ตรงและมีส่วนประกอบไม่ครบถ้วน ซึ่งแนวทางแก้ไขคือต้องถ่ายรูปใบหน้าสุนัขใหม่เท่านั้น
- 4) Android Studio ประมวลผลช้าและใช้เวลาในการทดสอบด้วย emulator เป็นเวลานาน แนวทางแก้ไขที่เป็นไปได้คือการทำงานบน cross-platform framework เช่น Ionic Xamarin หรือ Phonegap ซึ่งมีความรวดเร็วกว่าเนื่องจากทำงานบนเว็บเทคโนโลยี แต่ต้องคำนึงถึงเรื่องของการใช้งาน API ระดับลึก

5.3 แนวทางการพัฒนาต่อ

พัฒนาแอปพลิเคชันให้มีประสิทธิภาพมากขึ้น ทั้งในด้านของประสิทธิภาพ ความรวดเร็วในการประมวลผล ความแม่นยำในการค้นหาสุนัขที่มีลักษณะใบหน้าคล้ายกัน ส่วนติดต่อผู้ใช้งาน และประสบการณ์การใช้งานของผู้ใช้ เพื่อยกระดับให้สามารถกลายเป็นแอปพลิเคชันสำหรับชุมชนอาสาสมัครดูแลสุนัขได้อย่างครบวงจรในอนาคต

บรรณานุกรม

Thierry Pinheiro Moreira, Mauricio Lisboa Perez, Rafael de Oliveira Werneck, Eduardo Valle.

2016. **Where Is My Puppy? Retrieving Lost Dogs by Facial Features.**

ผศ.น.สพ.ปานเทพ รัตนากร. เมื่อหมาหาย. [Online]. Available : <https://goo.gl/Vpz4F3>

Cross Validated. 2010. Using principal component analysis (PCA) for feature selection . [Online]. Available : <https://goo.gl/nat3ly>

Wikipedia Contributors. 2017. **F1 Score.** [Online]. Available : <https://goo.gl/Ry8IGd>

Wikipedia Contributors. 2017. **Supervised learning.** [Online]. Available : <https://goo.gl/hPkcFR>

Wikipedia Contributors. 2015. **k-means clustering.** [Online]. Available : <https://goo.gl/MF40IH>

Wikipedia Contributors. 2015. **ระยะทางแบบยุคลิด.** [Online]. Available : <https://goo.gl/fxV0dR>

Wikipedia Contributors. 2017. **Affinity Propagation.** [Online]. Available : <https://goo.gl/KRk2rN>

MathWorks. 2017. **sharpen.** [Online]. Available : <https://goo.gl/SSmXTI>

คณะทรัพยากรธรรมชาติ มหาวิทยาลัยสงขลานครินทร์. **การปรับปรุงภาพ.** [Online].

Available : <https://goo.gl/Qp9i3S>

scikit.learn developers. 2010. **Euclidean Distances.** [Online].

Available : <https://goo.gl/xSWHzj>

scikit.learn developers. 2010. **Principal component analysis (PCA).** [Online].

Available : <https://goo.gl/GuvuaK>

scikit.learn developers. 2016. **K-means.** [Online]. Available : <https://goo.gl/jUNxaa>

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้