

ตัวแบบการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูล
ที่ไม่สมดุลในการพัฒนาระบบบริการตนเอง สำหรับศูนย์
นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

มหาวิทยาลัยเชียงใหม่

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568 **มกราคม 2567**

ตัวแบบการจำแนกข้อความแบบหลายคลาสตัวชุดข้อมูลใหม่
สมดุลในการพัฒนาระบบบริการตนเอง สำหรับศูนย์นวัตกรรม
การสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย

ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

มกราคม 2567

ตัวแบบการจำแนกข้อความแบบห้องคลาสด้วยชุดข้อมูลที่ไม่
สมดุลในการพัฒนาระบบบริการตนเอง สำหรับศูนย์นวัตกรรม
การสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่



การค้นคว้าอิสระนี้เสนอต่อมหาวิทยาลัยเชียงใหม่เพื่อเป็นส่วนหนึ่งของการศึกษา^๑
ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีและสาขาวิชาการจัดการ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright[©] by Chiang Mai University
All rights reserved

มหาวิทยาลัยเชียงใหม่

มกราคม 2567

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ตัวแบบการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่ไม่สมดุลในการ
พัฒนาระบบบริการตนเอง สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้
มหาวิทยาลัยเชียงใหม่

ปัณฑิ พันธุ์วุฒิ

การค้นคว้าอิสระนี้ได้รับการพิจารณาอนุมัติให้นำเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตร์มหาบัณฑิต
สาขาวิชาเทคโนโลยีและสหวิทยาการจัดการ

คณะกรรมการสอน

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สายฝน อุ่นนันกาศ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.เสนອแก้ว สมหอม)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ดุษฎี ประเสริฐธิพงษ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.วิจักPN ศรีสัจจะเลิศวิชา)

คณะกรรมการที่ปรึกษา

..... อาจารย์ที่ปรึกษาหลัก
(ผู้ช่วยศาสตราจารย์ ดร.ดุษฎี ประเสริฐธิพงษ์)

..... อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร.วิจักPN ศรีสัจจะเลิศวิชา)

19 มกราคม 2567

© ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

กิตติกรรมประกาศ

การค้นคว้าอิสระเรื่อง ตัวแบบการจำแนกข้อความแบบหดยคลาสคำวายชุดข้อมูลที่ไม่สมดุล
ในการพัฒนาระบบบริการตนเอง สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้
มหาวิทยาลัยเชียงใหม่ ได้รับความอนุเคราะห์จาก ผู้ช่วยศาสตราจารย์ ดร.ดุษฎี ประเสริฐธิพงษ์ รอง
ผู้อำนวยการสำนักทะเบียนและประมวลผล มหาวิทยาลัยเชียงใหม่ และผู้ช่วยศาสตราจารย์ ดร.
วิจักษณ์ ศรีสัจจะเลิศวิชา ผู้ช่วยศาสตราจารย์ ดร.สุริยา โนโลยสารสนเทศ คณะวิทยาศาสตร์
มหาวิทยาลัยเชียงใหม่ ที่ได้ให้ความกรุณาอนุมัติปริญญา ดำเนินการ ปรับปรุงแก้ไข ตลอดจนความรู้ที่
เกี่ยวข้องกับหัวข้อการค้นคว้า ล่างผล ทำให้การค้นคว้าประสบความสำเร็จ ได้อย่างราบรื่น

ขอขอบคุณ อาจารย์ ดร.อาณันท์ สีห์พิทักษ์เกียรติ ผู้อำนวยการศูนย์นวัตกรรมการสอนและการเรียนรู้
มหาวิทยาลัยเชียงใหม่ ที่ให้ความอนุเคราะห์ใช้ข้อมูลภายในศูนย์ฯ เพื่อการค้นคว้าในครั้งนี้

ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.เสมอແບ สมหอม และ อาจารย์ ดร. ดำรงศักดิ์ นภารัตน์
รองผู้อำนวยการสำนักบริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยเชียงใหม่ ที่ให้ความอนุเคราะห์ให้
คำปรึกษา แนวคิดอันเกี่ยวข้องกับหัวข้อค้นคว้าอันเป็นประโยชน์

ผู้ค้นคว้าคาดหวังว่าการค้นคว้านี้จะสามารถเป็นประโยชน์สำหรับหน่วยงานที่เกี่ยวข้อง และ
ผู้ที่สนใจศึกษาและต่อยอดต่อไป

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ปณิธ พันธุ์วุฒิ

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

หัวข้อการค้นคว้าอิสระ ตัวแบบการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่ไม่สมดุล
ในการพัฒนาระบบบริการต้นของ สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่

ผู้เขียน นาย ปณิธ พันธุ์วุฒิ

ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีและสาขาวิชาการจัดการ)

คณะกรรมการที่ปรึกษา ผศ.ดร.คุณณิ ประเสริฐนิติพงษ์ อาจารย์ที่ปรึกษาหลัก
ผศ.ดร.วิจักษณ์ ศรีสัจจะเลิศวاجา อาจารย์ที่ปรึกษาร่วม

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่ไม่สมดุลในการพัฒนาระบบบริการต้นของ สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ โดยดำเนินการตามกระบวนการเรียนรู้ของเครื่อง ได้แก่ การเตรียมข้อมูล การฝึกตัวแบบ การประเมินตัวแบบ และการตรวจสอบตัวแบบ โดยใช้ตัวแบบต่าง ๆ เพื่อประเมินประสิทธิภาพและความแม่นยำในการแก้ไขปัญหาของศูนย์นวัตกรรมการสอนและการเรียนรู้ ผลการวิจัยแสดงว่าการใช้ TF-IDF ร่วมกับ Multi-Layer Perceptron เป็นกระบวนการที่มีความแม่นยำที่สุดสำหรับชุดข้อมูลภาษาไทยที่ไม่สมดุลในการจำแนกข้อความนี้

งานวิจัยนี้เน้นความสำคัญของการเลือกตัวแบบที่เหมาะสมในการแก้ไขปัญหาการจำแนกข้อความ โดยเฉพาะอย่างยิ่งเมื่อมีการจัดการกับชุดข้อมูลที่ไม่สมดุล ตัวแบบที่ได้ที่พัฒนาขึ้นสามารถนำไปใช้เป็นกรอบการทำงานสำหรับระบบบริการต้นของในสถาบันการศึกษาอื่น ๆ เพื่อเพิ่มประสิทธิภาพการใช้งานระบบโดยรวม และผลการวิจัยนี้สามารถนำไปเป็นแนวทางสำหรับการพัฒนาตัวแบบการจำแนกข้อความในภาษาไทยต่อไป

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

Independent Study Title Multi-Class Text Classification Models from Imbalanced Datasets
for Self-Service System Development in Teaching and Learning
Innovation Center in Chiang Mai University

Author Mr. Panithi Panwut

Degree Master of Science (Technology and Interdisciplinary Management)

Advisory Committee Asst. Prof. Dr.Dussadee Praserttitipong Advisor
Asst. Prof. Dr.Wijak Srisujjalertwaja Co-advisor

ABSTRACT

The object of this study is to develop a self-service system for the Teaching and Learning Innovation Center at Chiang Mai University using multi-class text classification technique. This study employed an imbalanced Thai dataset and followed the standard machine learning pipeline of data preparation, model training, model evaluation, and review model. Different models were utilized to evaluate their performance and accuracy in solving the problem at hand. The results findings indicate that the use of TF-IDF with Multi-Layer Perceptron produced the most accurate and effective results for their dataset.

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

สารบัญ

หน้า

กิตติกรรมประกาศ	๑
บทคัดย่อภาษาไทย	๑
ABSTRACT	๑
สารบัญตาราง	๗
สารบัญรูปภาพ	๘
 บทที่ ๑ บทนำ	
1.1 ความเป็นมาและความสำคัญของปัญหา	๑
1.2 วัตถุประสงค์	๒
1.3 ขอบเขตของการวิจัย	๓
1.4 ประโยชน์ที่คาดว่าจะได้รับ	๗
 บทที่ ๒ ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	
2.1 ระบบบริการตนเอง	๘
2.2 การจำแนกข้อความ	๘
2.3 เทคนิคการสังเคราะห์ข้อมูลเพิ่มเติม	๑๑
2.4 การประเมินประสิทธิภาพ	๑๓
2.5 การนำไปใช้	๑๓
 บทที่ ๓ วิธีการดำเนินการ	
3.1 การเตรียมข้อมูล	๑๖
3.2 การฝึกตัวแบบ	๑๘
3.3 การประเมินตัวแบบ	๑๙
 บทที่ ๔ ผลการทดลองและวิจารณ์ผลการทดลอง	
4.1 ผลการทดสอบความแม่นยำ	๒๐

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริพัทธ์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

4.2	ผลการทดสอบความเที่ยงตรง	21
4.3	ผลการทดสอบค่าความครบถ้วน	22
4.4	ผลการทดสอบคะแนน F1	24
4.5	วิจารณ์ผลการทดลอง	25
บทที่ 5	สรุปผลการทดลอง	26
5.1	สรุปภาพรวมของการทดลอง	26
5.2	ข้อเสนอแนะ	26
5.3	ปัญหาในการวิจัย	27
เอกสารอ้างอิง		28
ภาคผนวก		
ภาคผนวก ก	ตัวอย่างของกระบวนการเตรียมข้อมูล	34
ภาคผนวก ข	เบริยมเทียบระยะเวลาก่อน-หลังการประมวลผลโดยใช้หน่วยประมวลผลกราฟิกของ CatBoost	36
ภาคผนวก ค	เบริยมเทียบผลการทดสอบก่อน-หลังการอัปเดตไลบรารี	37
ประวัติผู้เขียน		39

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
 Copyright[©] by Chiang Mai University
 All rights reserved

สารบัญตาราง

หน้า

ตารางที่ 1.1 คำอธิบายคลาสของข้อความที่ถูกจำแนกโดยผู้เชี่ยวชาญ	3
ตารางที่ 1.2 ตัวอย่างข้อความ	4
ตารางที่ 1.3 แสดงตัวอย่างข้อความที่ถูกตัดออก	6
ตารางที่ 3.1 การประมวลผลข้อมูล	16
ตารางที่ 3.2 อัลกอริทึมและพารามิเตอร์	18
ตารางที่ 4.1 ผลการทดสอบความแม่นยำ	20
ตารางที่ 4.2 ผลการทดสอบความเที่ยงตรง	21
ตารางที่ 4.3 ผลการทดสอบค่าความครบถ้วน	22
ตารางที่ 4.4 ผลการทดสอบคะแนน F1	24
ตารางที่ ก.1 ตารางเครื่อมข้อมูล	34
ตารางที่ ข.1 ตารางระยะเวลา ก่อน-หลังการประมวลผลโดยใช้หน่วยประมวลผลกราฟิก	36
ตารางที่ ค.1 ตารางผลการทดสอบก่อนอัปเดต ไลบรารี	37
ตารางที่ ค.2 ตารางผลการทดสอบหลังอัปเดต ไลบรารี	38

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

สารบัญรูปภาพ

หน้า

ภาพที่ 3.1	กระบวนการทดลอง	15
ภาพที่ 3.2	จำนวนข้อความที่ไม่เท่ากันในแต่ละคลาส	17
ภาพที่ 3.3	แสดงตัวอย่างจำนวนข้อความหลังจากใช้เทคนิค SMOTE	18



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

บทที่ 1

บทนำ

ในส่วนของบทนำนี้จะอธิบายถึงความเป็นมา สาเหตุและความสำคัญของปัญหานางาน
บริการการเรียนรู้ในรูปแบบดิจิทัล รวมถึงวัตถุประสงค์หลัก ขอบเขต และประโยชน์ที่ได้รับจากการ
นำตัวแบบการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่ไม่สมดุล ในการพัฒนาระบบบริการ
ต้นของสำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่

1.1 ความเป็นมาและความสำคัญของปัญหา

การพัฒนาระบบสนับสนุนบริการด้วยตัวเอง ได้รับความสนใจอย่างมากในการให้บริการที่
เกี่ยวข้องกับเทคโนโลยีสารสนเทศ ระบบเหล่านี้มีเป้าหมายในการสนับสนุนการให้บริการต้นของ
แบบอัตโนมัติที่มีประสิทธิภาพและมีประสิทธิผล อย่างไรก็ตามอุปสรรคสำคัญหนึ่งในการสร้าง
ระบบเช่นนี้คือ การจำแนกประเภทของข้อมูลที่ผู้ใช้ป้อนเข้าระบบให้แม่นยำ โดยเฉพาะอย่างยิ่งใน
บริบทของการจำแนกประเภทของข้อความแบบหลายคลาส (Multi-class Text Classification) อีก
ทั้งหากมีชุดข้อมูลที่ไม่สมดุลก็จะเพิ่มความซับซ้อนให้กับงาน ด้วยเหตุนี้การแก้ไขข้อมูลที่ไม่สมดุล
ในการจำแนกประเภทของข้อความแบบหลายคลาสเป็นสิ่งสำคัญสำหรับการพัฒนาระบบสนับสนุน
บริการด้วยตัวเอง (Revina et al., 2020)

การจำแนกประเภทข้อความแบบหลายคลาสใช้เพื่อจัดประเภทข้อมูลของข้อความหลาย
ประเภท สิ่งสำคัญคือต้องสังเกตว่าประสิทธิภาพของอัลгорิทึมที่ใช้ขึ้นอยู่กับลักษณะของข้อมูลที่เป็น^{ลักษณะของมหาวิทยาลัยเชียงใหม่ ที่อยู่ในประเทศไทย จังหวัดเชียงใหม่}
ข้อความนั้น การจำแนกประเภทข้อความแบบหลายคลาสมักใช้กับข้อมูลที่ไม่สมดุล (Imbalance Dataset)
ซึ่งหมายความว่าบางคลาสมีข้อมูลตัวอย่างมากกว่าคลาสอื่นอย่างมีนัยสำคัญ (Raza et al.,
2019)

ปัญหาของชุดข้อมูลที่ไม่สมดุลนี้สามารถดำเนินไปสู่แบบจำลองที่มีอคติซึ่งทำงานได้ไม่ดีใน
คลาสที่มีข้อมูลตัวอย่างน้อย ผลที่ตามมาของข้อมูลที่ไม่สมดุล ได้แก่ ความแม่นยำ (Accuracy)
และความเที่ยงตรง (Precision) ที่ลดลง ทั้งนี้การจัดการกับชุดข้อมูลที่ไม่สมดุลด้วยเทคนิคการสังเคราะห์
ข้อมูลเพิ่มเติม (Synthetic Minority Over-sampling Technique : SMOTE) สามารถปรับปรุง
ประสิทธิภาพของตัวแบบการจำแนกประเภทของข้อความในคลาสที่มีข้อมูลตัวอย่างน้อยด้วยการ

สร้างตัวอย่างสังเคราะห์ เพื่อช่วยปรับสมดุลข้อมูลและลดค่าติดต่อคลาสส่วนใหญ่ได้ เป็นเทคนิคที่มีประโยชน์สำหรับจัดการกับข้อมูลที่ไม่สมดุลในการจำแนกประเภทของข้อความแบบหลายคลาส และสามารถช่วยปรับปรุงความแม่นยำ และความเที่ยงตรง ของตัวแบบในคลาสที่มีข้อมูลตัวอย่างน้อยกว่า (Batista et al., 2004 and Raza et al., 2019)

ในการศึกษานี้ผู้ค้นคว้านำเสนอความก้าวหน้าของการพัฒนาระบบจำแนกประเภท ข้อความแบบหลายคลาสสำหรับระบบสนับสนุนการให้บริการด้วยตัวเองในบริบทของศูนย์นวัตกรรมการสอนและการเรียนรู้ ของมหาวิทยาลัยเชียงใหม่ ที่มีปัญหาการจำแนกประเภทการสนทนาระบบข้อความในระบบสนับสนุนการให้บริการ ที่มีชุดข้อมูลตัวอย่างที่ไม่สมดุล โดยมีงานวิจัยที่มีอยู่แล้วในสาขาวิชาของการเรียนรู้ของเครื่อง (Machine Learning) การจำแนกประเภทของข้อความ (Text Classification) โดยผู้ค้นคว้าจะนำเสนอแนวทางใหม่ที่ใช้เทคนิคที่กล่าวข้างต้น รวมถึงการประมวลผลธรรมชาติภาษา เพื่อนำการจำแนกข้อความแบบหลายคลาスマใช้ในบริบทของการให้บริการเฉพาะของศูนย์นวัตกรรมการสอนและการเรียนรู้ โดยคาดหวังว่าการค้นคว้านี้จะมีศักยภาพในการปรับปรุงคุณภาพการให้บริการของศูนย์นวัตกรรมการสอนและการเรียนรู้ ของมหาวิทยาลัยเชียงใหม่อย่างมีนัยสำคัญ โดยการจำแนกประเภทข้อมูลที่ป้อนเข้ามาจากผู้ใช้อย่างแม่นยำ และมีประสิทธิภาพ มีศักยภาพในการตอบสนองต่อความท้าทายที่เกิดขึ้นจากชุดข้อมูลภาษาไทยที่ไม่สมดุล และมีลักษณะที่หลากหลาย เช่น ภาษาท้องถิ่น ภาษาต่างประเทศ ฯลฯ ที่มีความซับซ้อน แต่สามารถจัดการได้ด้วยความสามารถของเครื่องจักร ทำให้การจำแนกประเภทของข้อความในบริบทของการสอนและการเรียนรู้ ของมหาวิทยาลัยเชียงใหม่

1.2 วัตถุประสงค์

เพื่อพัฒนาตัวแบบการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่ไม่สมดุลในการพัฒนาระบบบริการตนเอง สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่

1.3 ขอบเขตของการวิจัย

1.3.1 ประชากรและตัวอย่างวิจัย

การทดลองนี้ได้เก็บรวบรวมข้อมูลความการติดต่อผ่านทางหน้าจอช่วยเหลือของระบบจัดการการเรียนรู้ ที่ศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ จัดเก็บ ซึ่งเป็นชุดข้อมูลที่มีจำนวนข้อมูลตัวอย่างในแต่ละคลาส โดยถูกกำหนดขึ้นจากประเภทของปัญหาที่ผู้ใช้ได้ทำการสอบถามเข้ามายังหน้าจอช่วยเหลือ ระหว่างเดือน มีนาคม 2564 - มีนาคม 2565 จำนวน 1,000 ข้อความ โดยใช้แพลตฟอร์มที่เกี่ยวข้องกับระบบบริหารการจัดการเรียนรู้เนื่องจากเป็นบริการที่มีข้อมูลต่อคลาสสูงที่สุด สามารถแบ่งออกเป็น 8 คลาสหลัก ซึ่งคำอธิบายคลาสทั้ง 8 คลาสถูกระบุไว้กับเกณฑ์การจำแนกดังตารางที่ 1.1

ตารางที่ 1.1 : แสดงคำอธิบายคลาสของข้อมูลที่ถูกจำแนกโดยผู้ใช้ชาวญี่ปุ่น

คลาส (Class)	เกี่ยวข้องกับ	เกณฑ์จำแนก
Account	บัญชีผู้ใช้	ปัญหาการเข้าใช้งาน เช่น ล็อกอินไม่ผ่านหรือลืมรหัสผ่าน
Activity	กิจกรรมในวิชา	คำตามที่เกี่ยวข้องกับการใช้กิจกรรมในกระบวนการวิชาออนไลน์
Course Portal	การเปิดใช้วิชา	ความต้องการที่จะเปิดใช้งานกระบวนการวิชาออนไลน์
Course Setting	การตั้งค่าในวิชา	ความต้องการที่จะปรับแต่งการตั้งค่าของวิชา
Enrollment	การเข้าถึงวิชา	ความต้องการที่ผู้ใช้ต้องการเข้าดูเนื้อหาในกระบวนการวิชา
Grade	ผลการเรียน	คำตามเกี่ยวกับคะแนนในวิชา

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายธีรวิทย์ อันทรงเพย์
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ตารางที่ 1.1 : แสดงคำอธิบายคลาสของข้อความที่ถูกจำแนกโดยผู้เชี่ยวชาญ (ต่อ)

คลาส (Class)	เกี่ยวข้องกับ	จำแนกจาก
Quiz	ข้อสอบ	คำตามเกี่ยวกับการใช้งาน หรือตั้งค่าที่เกี่ยวข้องกับการสอน
Other	ส่วนที่นอกเหนือหัวข้อที่ให้บริการ	คำตามอื่นๆที่ไม่เกี่ยวข้องกับหัวข้อที่ให้บริการ

ในชุดข้อมูลขนาดเล็กที่ไม่สมดุล คลาสส่วนน้อยจะแสดงผลได้ไม่คุ้มค่าจำนวนข้อมูลตัวอย่างที่ลดลงมากเกินไปซึ่งอาจไม่เพียงพอสำหรับให้เครื่องเรียนรู้ทั้งนี้ยังไม่มีการกำหนดข้อมูลจำนวนน้อยที่สุดที่สามารถฝึกฝนในชุดข้อมูลที่ไม่สมดุลได้เนื่องจากขึ้นอยู่กับขนาดและความซับซ้อนของชุดข้อมูล ซึ่งหลังจากผู้เชี่ยวชาญได้ทำการระบุคลาสในกลุ่มตัวอย่างแล้วสามารถคัดเลือกข้อความที่适合คลาสส่วนต้นได้จำนวน 529 ข้อความ ซึ่งแสดงตัวอย่างชุดข้อมูลตัวอย่างของแต่ละคลาสได้ดังตารางที่ 1.2

ตารางที่ 1.2 : แสดงตัวอย่างข้อความ

ข้อความ	คลาส
ไม่สามารถเข้าสู่ระบบได้ หนูไม่สามารถเข้าสู่ระบบได้ แก้ไขให้หน่อยค่ะ !IMG_0313.jpg	Account
Log-in cmu account ไม่ได้ ถ้าเข้าจากอุปกรณ์อื่น สามารถเข้าได้ แต่เข้าจาก chrome ในโทรศัพท์แล้วมีปัญหาครับ ควรแก้หรือรีเซ็ตอะไรยังไงครับ	Account
สอบถามวิธีการให้นักศึกษาส่งไฟล์งานในระบบ KC Moodle xxx xxx ติดต่อเรื่อง การให้นักศึกษาส่งไฟล์งานในระบบ KC Moodle	Activity

ตารางที่ 1.2 : แสดงตัวอย่างข้อความ (ต่อ)

ข้อความ	คลาส
สอบถามวิธีส่งงาน kc moodle ค่ะ สอบถามค่ะ จากนักศึกษาค่ะ พอดีอาจารย์ กำหนดให้ส่งงานโดย เรียงไฟล์เป็นลำดับชื่อไฟล์ตามที่อาจารย์กำหนด ซึ่ง ตอนดึงไฟล์มาอัพโหลดลง ก็ดึงตามที่อาจารย์กำหนด แต่พอโหลดที่เรามันไม่ ตรงกับที่อาจารย์สั่ง มีวิธีแก้ไขใหม่ค่ะ ต้องคลิกตรงไหนเพิ่มเติมใหม่ค่ะ	Activity
รบกวนขอเปิดกระบวนการวิชาหน่อยครับ รบกวนขอเปิดกระบวนการวิชา xxxxxxx บน moodle ด้วยครับ ขอบคุณครับ	Course Portal
ขอเพิ่มกระบวนการวิชา xxxxxxx ใน kc moodle ค่ะ เรียนท่านผู้เกี่ยวข้อง อาจารย์ จะขอเพิ่มกระบวนการวิชา xxxxxxx ลงใน kc moodle ขอบคุณมากค่ะ อ xxxxxxx (xxxxxx) !Outlook-signature3.png Course Portal	Course Portal
ลบ KC Moodle เรียนทีมที่ลิค พอดีลองกดสร้างคอร์สค่ะ แล้วมันซ้ำกัน [xxx-xxx: Participants (cmu.ac.th) https://elearning.cmu.ac.th/user/xxx] [elearning.cmu.ac.th/user/xxx http://elearning.cmu.ac.th/user/xxx] รบกวนลบ KC Moodle ก้อนนี้ทิ้งที่ค่าพอดีสร้างซ้ำค่ะ อันที่ใช้จริงคืออีกอัน ค่ะ	Course Setting
ต้องการเปลี่ยนรูปหน้าปกวิชาใน Dashboard ขอทราบวิธีการเปลี่ยนรูปหน้าปกวิชาใน Dashboard	Course Setting
19 ยังไม่ถูกดึงเข้าวิชา เป็นเรียนไม่ได้ ทำอย่างไร? (7:15:49 AM)	Enrollment
12 สวัสดีค่ะ (9:24:58 AM) หนูเพิ่ม kc moodle xxx ไม่ได้ค่ะ (9:25:28 AM) t1.15752-9/xKA62F404C4 secxxx xxx xxx@cmu.ac.th ค่ะ (12:24:12 PM) นายวิริวิทย์ จันทร์เพย ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568	Enrollment

ตารางที่ 1.2 : แสดงตัวอย่างข้อความ (ต่อ)

ข้อความ	คลาส
วิธีใช้ CMU grade report เรียน ผู้ที่เกี่ยวข้อง ระบบขอวิธีใช้ grade report สำหรับประกาศคะแนนสอบนักศึกษารายบุคคล ในโปรแกรม CMU Online ด้วยครับ ขอบคุณครับ -- !https://xxx xxx	Grade
ระบบ KC moodle ไม่มีการแสดงผลคะแนน ในรายวิชา 715 Quiz topic ที่ 3 ระบบ KC moodle ไม่มีการแสดงผลคะแนน ต้องทำอย่างไรบ้างคะ	Grade
การแก้ไขผลคำตอบและให้ระบบตรวจใหม่ทำไงครับ	Quiz
อีกคำถามค่ะ ถ้าแก้ไขลงในข้อสอบchoice หลังจากที่ นศ ทำข้อสอบเสร็จไปแล้ว จะ process คะแนนให้ นศ ใหม่ต้องทำอย่างไรค่ะ	Quiz
17 มีการบริการรับลงทะเบียนโปรแกรม SPSS ใหม่ๆหรือพอจะสามารถหาลงได้ที่ไหนคะ (3:01:44 AM) ขอบคุณค่ะ (3:25:13 AM) Install safe exam browser ไม่สำเร็จ Install แอพ safe exam ไม่ได้ค่ะ - ได้ลองโหลดแอพเพิ่มตามที่เขียนบอกแล้วค่ะคือ .net framework 4.7.2 และ Visual c++ ก็ปัง install ไม่ได้อยู่ดีค่ะ	Other

หลังจากคัดเลือกข้อความ ได้ 529 ข้อความข้างต้นแล้วตารางที่ 1.3 แสดงตัวอย่างข้อความที่ถูกตัดออก
จำนวน 471 ข้อความ ดังนี้

ตารางที่ 1.3 : แสดงตัวอย่างข้อความที่ถูกตัดออก

ข้อความ	เหตุผล
ขาดสอบวิชา [REDACTED] ผนท้องเสียและปวดท้องหนักมากๆตั้งแต่เช้าวันที่ 11 และช่วง 11 โมงคุณพ่อพากจากที่ทำงานจึงมารับผนทีบ้านพาไปโรงพยาบาล ซึ่งตอนนั้นผนทีไข้และอาเจียนอ่อนเพลียมากๆครับ พอนำไปถึงโรงพยาบาล เนื่องจากผนทีอาการอ่อนเพลียคุณหมอมึงให้นอนดูอาการในห้องฉุกเฉิน และทำการเติมน้ำเกลือและนีดยาแต่คนไข้ค่อนข้างเยอะจึงต้องรอคิวรับจึงทำให้ผนทีไม่สามารถเข้าสอบได้ทันเวลาครับผนท. และผนท. ไม่มีอินเตอร์เน็ต ด้วยครับหลังจากนี้ด้วยแล้วอาการผนท. ไม่ดีขึ้นเท่าไหร่นักคุณหมอมึงให้ นอนดูอาการในห้องฉุกเฉินจนกว่าจะดีขึ้น เพราะหากให้ผนท.กลับบ้านอาจเกิดการซื้อกินมาได้และผนท. แจ้งความต้องการไม่นอนโรงพยาบาลครับ	เป็นข้อมูลทั่วไปที่ไม่เกี่ยวข้องกับการให้บริการใดๆ
Excel SUM ไม่ได้ ทำการรวมแบบกด sum ไม่ได้ แต่ใส่สูตรทำได้ ใน Column AN ครับ	เป็นคำตามที่อยู่นอกเหนือบริการที่ให้บริการ และไม่สามารถระบุรายละเอียดอื่นๆ ได้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ จะได้รับแบบจำลองที่เหมาะสมในการจำแนกปัญหาของผู้ใช้งาน

1.4.2 ศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ สามารถนำแบบจำลองที่ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริพัฒน์ จันทร์เทย ดาวน์โหลดเมื่อ [REDACTED] ได้รับมาดำเนินทดสอบสนับสนุนผู้ใช้บริการในส่วนงานสนับสนุนการบริการของศูนย์ฯ

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในปัจจุบันธุรกิจหลากหลายประเภทและองค์กรต่าง ๆ ให้ความสนใจกับการนำเทคโนโลยีสารสนเทศเข้ามาช่วยสนับสนุนในการบริหารจัดการทางด้านไอที เพื่อเพิ่มขีดจำกัดและพัฒนาศักยภาพการให้บริการ นอกเหนือจากนั้นยังสามารถสร้างภาพลักษณ์ขององค์กรให้มีมาตรฐานที่ดีขึ้น เพื่อตอบสนองความต้องการ และยกระดับงานบริการทางด้านไอทีให้มีประสิทธิภาพที่สูงขึ้น จึงได้นำระบบบริการตนเองมาใช้เพื่อให้สามารถตอบสนองความต้องการของผู้ใช้บริการได้อย่างทันท่วงที

2.1 ระบบบริการตนเอง

ระบบบริการตนเอง (Self-service System) เป็นแพลตฟอร์มหรือส่วนต่อประสานทางเทคโนโลยี (Technological Interface) สำหรับโต้ตอบระหว่างผู้ใช้กับตัวแทนขององค์กร ช่วยให้ผู้ใช้สามารถดำเนินการในส่วนของการบริการผ่านส่วนต่อประสานได้ด้วยตัวเอง มีเป้าหมายเพื่อส่งเสริมประสิทธิภาพและคุณภาพบริการที่ดี การค้นหาข้อมูลทางอินเทอร์เน็ตเป็นตัวอย่างหนึ่งของ การบริการตนเองที่ช่วยให้ผู้ใช้บริการได้รับข้อมูลและช่วยเหลือตนเอง ซึ่งสามารถในค้นหาข้อมูลได้อย่างรวดเร็ว และพึงพอใจกับการค้นหาเมื่อพบข้อมูลที่ต้องการ โดยประยุกต์ของการบริการตนเอง แบบออนไลน์อาจรวมถึงความสะดวกสบาย ประสิทธิภาพ และความยืดหยุ่น ได้รับความประทับใจในการเข้าถึงบริการได้ทุกที่ทุกเวลาทางออนไลน์ (Meuter et al., 2000) สามารถปรับปรุงความแม่นยำของการกำหนดเส้นทางการค้นหา ด้วยการใช้การจำแนกประเภทข้อความแบบหลายคลาส (Multi-class Text Classification) และนำไปสู่การแก้ปัญหาของผู้ใช้งานได้เร็วขึ้น

2.2 การจำแนกข้อความ

การจำแนกข้อความ (Text Classification) เป็นเทคนิคนึงในการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) ที่เกี่ยวข้องกับการจำแนกประเภทข้อความ หรือหมวดหมู่ที่กำหนดไว้ล่วงหน้าหลายรายการ โดยหนึ่งในเทคนิคสำหรับการจำแนกประเภท ข้อความแบบหลายคลาส ก็คือ แมชชีนเลิร์นนิงอัลกอริทึม (Machine Learning Algorithms) ที่ลิสต์ที่อธิบายประกอบด้วยอัลกอริทึมต่างๆ เช่น ชัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine : SVM) ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568 เนอีฟเบย์ (Naïve Bayes) และต้นไม้การตัดสินใจ (Decision Trees) ซึ่งเป็นเทคนิคที่ได้รับการวิจัยว่าทำงานได้ดีในการจำแนกประเภทข้อความแบบหลายคลาส (Zhang and Zhou, 2015)

2.2.1 ชั้พพอร์ตเวคเตอร์แมชชีน

ชั้พพอร์ตเวคเตอร์แมชชีน (Support Vector Machines : SVM) เป็นเทคนิคที่ใช้กันทั่วไปสำหรับงานจำแนกข้อความ เนื่องจากสามารถจัดการความสัมพันธ์ทึ้งเชิงเส้นและไม่เชิงเส้นระหว่างคุณลักษณะและคลาสได้อย่างมีประสิทธิภาพ

ในการจำแนกข้อความแบบหลายคลาส ชั้พพอร์ตเวคเตอร์แมชชีนถูกใช้เพื่อเรียนรู้ของเขตการตัดสินใจที่แยกคลาส ตามคุณลักษณะต่างๆ ที่ดึงมาจากข้อความ คุณลักษณะสามารถแสดงเป็นเวกเตอร์ โดยที่แต่ละรายการสอดคล้องกับคำหรือคำเฉพาะในข้อความ มีจุดมุ่งหมายเพื่อแยกตัวอย่างเชิงบวกและเชิงลบออกจากพื้นที่คุณลักษณะ ของเขตการตัดสินใจถูกกำหนดโดยชุดย่อของตัวอย่าง การฝึกอบรมที่เรียกว่าเวกเตอร์สนับสนุน ซึ่งเป็นตัวอย่างที่ใกล้กับขอบเขตการตัดสินใจมากที่สุด (Joachims, 1998)

2.2.2 เนอีฟเบย์

เนอีฟเบย์ (Naive Bayes) เป็นเทคนิคที่ใช้งานง่าย มีความเร็วในการประมวลผล และใช้กับงานได้หลากหลาย (Rennie et al., 2003) จึงทำให้เป็นเทคนิคที่ใช้กันอย่างแพร่หลายในการจำแนกประเภทเอกสาร เนื่องจากมีประสิทธิภาพในการคำนวณและประสิทธิภาพในการทำนายที่ค่อนข้างดี (Frank & Bouckaert, 2006)

เป็นตัวจำแนกประเภทโดยใช้ความน่าจะเป็นอย่างง่ายซึ่งต้องยุ่บรวมติดกันที่ว่าคุณลักษณะทั้งหมดไม่เป็นอิสระจากกัน โดยพิจารณาจากตัวแปรหมวดหมู่ ซึ่งช่วยให้สามารถคำนวณได้อย่างมีประสิทธิภาพและมีเสถียรภาพ ทำให้เนอีฟเบย์เป็นตัวเลือกยอดนิยมในการจำแนกประเภทข้อความ (Xu, 2016)

2.2.3 การถดถอยโลจิสติก

การถดถอยโลจิสติก (Logistic Regression) เป็นรูปแบบการเรียนรู้ทางสถิติที่ใช้กันทั่วไปสำหรับงานจำแนกประเภท รวมถึงการจำแนกข้อความแบบหลายคลาส (Vlachos, 2008)

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ใบอนุญาตใช้งาน 27/11/2561-09:59:17 และหมดอายุ 06/05/2568
เนื่องจากมีประสิทธิภาพในการจัดการข้อมูลที่มีมิติสูง เช่น ข้อความภาษาธรรมชาติ ซึ่งมีประโยชน์

อย่างยิ่งในงานต่างๆ เช่น การวิเคราะห์ความรู้สึก การตรวจจับสแปม และการจัดหมวดหมู่หัวข้อ (Genkin et al., 2007) ซึ่งมักจะถูกเปรียบเทียบกับอัลกอริทึมการจำแนกประเภทอื่น ๆ เช่น ชัฟฟอร์ต เวคเตอร์แมชีน และต้นไม้การตัดสินใจ (Eichelberger & Sheng, 2013)

2.2.4 แรนดอมฟอร์ेसต์

แรนดอมฟอร์ेसต์ (Random Forest) เป็นเทคนิคซึ่งใช้สำหรับการจำแนกข้อความแบบหลายคลาส ที่ได้รับความนิยมเนื่องจากมีประสิทธิภาพที่โดดเด่นในงานจำแนกประเภท รวมถึงการจำแนกข้อความ มีการใช้สำหรับงานจำแนกข้อความต่างๆ รวมถึงการตรวจจับคำพูดแสดงความเกลียดชัง (Fauzi, 2018)

ในการจำแนกข้อความแบบหลายคลาส แรนดอมฟอร์ेसต์มักได้รับการเปรียบเทียบกับ แบบจำลองการทำนายทั่วไปอื่นๆ เช่น เนอฟเบย์ ชัฟฟอร์ต เวคเตอร์แมชีน และการตัดถอยโลจิสติก ซึ่งการเปรียบเทียบเหล่านี้แรนดอมฟอร์ेसต์มักจะได้ผลลัพธ์ที่มีประสิทธิภาพที่ดีในการจำแนก ข้อความ (Guan et al., 2019)

2.2.5 ต้นไม้การตัดสินใจ

ต้นไม้การตัดสินใจ (Decision Tree) เป็นเทคนิคการจำแนกประเภทที่ใช้แบบจำลองคล้ายต้นไม้ในการตัดสินใจ โดยพิจารณาจากคุณสมบัติของข้อมูลที่นำเข้า เป็นเทคนิคการเรียนรู้ที่สามารถใช้ได้กับปัญหาการจำแนกประเภทแบบหลายคลาส (Holmes et al., 2002) ในต้นไม้การตัดสินใจ แต่ละโหนดภายในแสดงถึงคุณลักษณะ และแต่ละกิ่งแสดงถึงค่าที่เป็นไปได้หรือผลลัพธ์ของคุณลักษณะนั้น ใบของต้นไม้เป็นตัวแทนของป้ายกำกับหรือคลาส

ในการจำแนกข้อความแบบหลายคลาส ข้อดีของแผนผังการตัดสินใจคือสามารถรับ ข้อมูลได้หลากหลายรูปแบบ เช่น ข้อมูลการจำแนกกลุ่ม ตัวเลข และข้อความ เป็นเทคนิคที่เข้ากันได้ดี และประมวลผลเร็วกับข้อมูลขนาดใหญ่ที่ไม่ต้องการการคำนวนที่ซับซ้อน (Rokach, 2016) ข้อดีอีกอย่างหนึ่งของแผนผังการตัดสินใจ คือมีความสามารถในการจัดการชุดข้อมูลที่ไม่สมดุล ซึ่งชุดข้อมูลที่ไม่สมดุลเป็นเรื่องปกติในงานจำแนกประเภทข้อความ เนื่องจากบางคลาสอาจมีข้อมูลน้ำเข้ามากกว่า

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่
คลาสอื่นๆ อย่างมาก แผนผังการตัดสินใจสามารถจัดการชุดข้อมูลที่ไม่สมดุลได้โดยการปรับเกณฑ์
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568
การตัดสินใจที่แต่ละโหนด ช่วยให้ประสิทธิภาพการจำแนกประเภทในคลาสที่มีจำนวนข้อมูลน้อยดี

ขึ้น (Alambo et al., 2020) และยังถูกนำมาใช้ในวิธีการแยกคำภาษาไทย ทำให้ได้รับความแม่นยำสูงกว่าเมื่อเปรียบเทียบกับตัวแยกประเภทอื่นๆ (Xiong et al., 2022)

2.2.6 แคทบูส

แคทบูส (CatBoost) เป็นเทคนิคที่มาจากการคำว่า "Categorical" และ "Boosting" เป็นเฟรมเวิร์ก Grading Boosting Decision Tree (GBDT) ที่เป็นการต่อ กันของต้นไม้การตัดสินใจหลายๆ ต้น ซึ่งมีชื่อเดียวกันในด้านประสิทธิภาพการจำแนกประเภทที่ยอดเยี่ยม (Prokhorenkova et al., 2017) เหมาะอย่างยิ่งสำหรับงานที่เกี่ยวข้องกับข้อมูลที่เป็นหมวดหมู่และข้อมูลที่แตกต่างกัน (Hancock & Khoshgoftaar, 2020)

ในการจำแนกข้อมูลที่ซับซ้อน เช่น การตรวจจับผู้เรียนที่ง่วงนอนในแพลตฟอร์มอีเลิร์นนิ่ง แคทบูส ได้รับการทดสอบและเปรียบเทียบกับชั้พพอร์ตเวคเตอร์แมชีน และแรนdomฟอร์สต์ ซึ่งได้ผลลัพธ์ที่มีประสิทธิภาพ (Kawamura et al., 2021)

2.2.7 เพอร์เซ็ปตรอนหลายชั้น

เพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron : MLP) เป็นเครือข่ายประสาทเทียมประเภทหนึ่งที่ใช้กันทั่วไปสำหรับการจำแนกข้อความแบบหลายคลาส (Athavale et al., 2019) ที่มีประสิทธิภาพอย่างยิ่ง เนื่องจากความสามารถในการสร้างฟังก์ชันการทำงานที่ซับซ้อน (Buabeng et al., 2021)

ในการจำแนกข้อความแบบหลายคลาส วิธีการเพอร์เซ็ปตรอนหลายชั้น แสดงให้เห็นว่ามีประสิทธิภาพเหนือกว่าวิธีการจำแนกด้วยต้นไม้การตัดสินใจ นอกจากนี้ การใช้กลยุทธ์การจำแนกประเภทแบบลำดับชั้นด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น ยังพบว่าช่วยเพิ่มความแม่นยำของการจำแนกประเภทได้อย่างมาก (Roeck et al., 2009) ซึ่งให้เห็นว่าวิธีการเพอร์เซ็ปตรอนหลายชั้น เหมาะอย่างยิ่งสำหรับการจัดการความซับซ้อนและความแตกต่างของงานการจำแนกข้อความแบบหลายคลาส

2.3 เทคนิคการสังเคราะห์ข้อมูลเพิ่มเติม

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ ในการจัดประชุมวิชาการที่มีหัวข้อ “สิ่งสำคัญคือต้องสร้างสมดุลของข้อมูลเพื่อบริหารจัดการความเรอนอียง ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568”
ของตัวแบบและปรับปรุงประสิทธิภาพในการจำแนกข้อความ เทคนิคการสังเคราะห์ข้อมูลเพิ่มเติม (Synthetic Minority Over-sampling Technique : SMOTE) เป็นเทคนิคสำหรับการปรับสมดุล

ชุดข้อมูลที่ไม่สมดุล สำหรับการจำแนกประเภทข้อความ โดยการสร้างตัวอย่างสังเคราะห์ของคลาส ย่อยขึ้นมา เพื่อปรับปรุงความแม่นยำในการจำแนกประเภทของคลาสย่อยขึ้น (Chawla, 2002) ซึ่งจากการศึกษาวิจัยของ Sun และคณะ (2017) และ Batista และคณะ (2018) พบว่าเป็นเทคนิคที่มีประสิทธิภาพสูงสุดในการปรับสมดุลชุดข้อมูลในการจำแนกข้อความ โดยจากการศึกษาวิจัยของ Chawla (2002) SMOTE มีหลักการทำงานดังนี้

1. สุ่มเลือกข้อความของคลาสที่มีจำนวนน้อย
2. หาจำนวนข้อมูลที่ใกล้เคียงจากข้อความที่เลือก
3. สุ่มเลือกหนึ่งในข้อมูลนั้น
4. สร้างข้อความใหม่โดยการประมาณค่าระหว่างข้อความต้นฉบับกับข้อมูลใกล้เคียงที่เลือก
5. ทำซ้ำขั้นตอนที่ 1-4 สำหรับตัวอย่างสังเคราะห์ตามจำนวนที่ต้องการ

ตัวอย่าง เช่น สมมติว่าเรามีกล่องใส่ลูกอมที่มีลักษณะต่างๆ อยู่ 100 เม็ด แต่ลูกอมสีชมพูมี 95 เม็ด และลูกอมสีฟ้ามี 5 เม็ด แบบนี้เรียกว่าชุดข้อมูลที่ไม่สมดุล (imbalanced dataset) เพราะลูกอมสีชมพูมากกว่าลูกอมสีฟ้ามาก ทำให้อัลกอริทึมการเรียนรู้ของเครื่อง เรียนรู้ลูกอมสีชมพูได้ดีกว่าลูกอมสีฟ้า ซึ่ง SMOTE คือเทคนิคที่ช่วยทำให้ชุดข้อมูลสมดุลขึ้น โดยการสร้างลูกอมสีฟ้าขึ้นมาใหม่ โดยใช้ข้อมูลของลูกอมสีฟ้าที่มีอยู่เดิม โดยมีขั้นตอนดังนี้

1. เลือกลูกอมสีฟ้าเม็ดหนึ่งขึ้นมา
2. หาลูกอมสีฟ้าที่อยู่ใกล้ลูกอมเม็ดแรกมากที่สุด
3. ยึดลักษณะบางอย่างจากลูกอมเม็ดที่สอง
4. สร้างลูกอมสีฟ้าเม็ดใหม่ขึ้นมาระหว่างลูกอมเม็ดแรกและเม็ดที่สอง ลูกอมเม็ดใหม่จะมีลักษณะคล้ายลูกอมเม็ดที่สอง แต่สียังคงเป็นสีฟ้า
5. ทำแบบนี้ซ้ำไปเรื่อย ๆ จนกว่าจะมีลูกอมสีฟ้าเพียงพอ

การสร้างลูกอมสีฟานี้จะช่วยทำให้ชุดข้อมูลสมดุลขึ้น เป็นการปรับสมดุลชุดข้อมูลทำให้อัลกอริทึมการเรียนรู้ของเครื่องสามารถเรียนรู้ลูกอมสีฟ้าได้ดีขึ้น

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริพัทธ์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

2.4 การประเมินประสิทธิภาพ

การประเมินประสิทธิภาพเป็นสิ่งสำคัญของแมชชีนเลร์นนิ่ง ซึ่งเป็นกระบวนการฝึกให้เครื่องเรียนรู้จากข้อมูลและทำการตัดสินใจ การวัดประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง เพื่อประเมินว่าเครื่องจะสามารถสรุปผลของข้อมูลที่รับเข้ามาใหม่ได้ดีเพียงใด โดยมีเมทริกซ์การประเมินหลายรายการเพื่อวัดประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง ได้แก่ ความแม่นยำ (Accuracy) ร้อยละของตัวอย่างที่จำแนกอย่างถูกต้อง เป็นอัตราส่วนของข้อมูลที่ทำนายได้อย่างถูกต้องต่อจำนวนข้อมูลทั้งหมด

$$\text{Accuracy} = \frac{\text{TN}+\text{TP}}{\text{TN}+\text{FP}+\text{TP}+\text{FN}} \quad (1)$$

ค่าความเที่ยงตรง (Precision) เป็นร้อยละของกลุ่มตัวอย่างที่ถูกจำแนกอย่างถูกต้องในกลุ่มตัวอย่างที่ได้คาดการณ์ไว้

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (2)$$

ค่าความครบถ้วน (Recall) เป็นร้อยละของตัวอย่างที่จำแนกอย่างถูกต้องในบรรดาตัวอย่างทั้งหมด

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (3)$$

และคะแนน F1 ที่เป็นค่าเฉลี่ยหารของนิภัยของความเที่ยงตรงและความครบถ้วน

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

2.5 การนำไปใช้

Alexander Genkin และคณะ (2007) ได้ทำการทดสอบอัลกอริทึมระหว่าง Logistic Regression กับ SVM ในการจัดหมวดหมู่ข้อความโดยใช้ข้อมูลข่าวสาร บทคัดย่อทางการแพทย์ และเนื้อหาจากหน้าเว็บไซต์ โดยข้อมูลถูกเปลี่ยนเป็นเวกเตอร์ด้วยเทคนิค TF-IDF ได้ผลลัพธ์ที่ Logistic Regression ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายชูรุทัย จันทร์พงษ์ หน้าเว็บไซต์ โดยข้อมูลถูกเปลี่ยนเป็นเวกเตอร์ด้วยเทคนิค TF-IDF ได้ผลลัพธ์ที่ Logistic Regression ด้านโน๊ตเดียว กับ SVM ได้ผลลัพธ์ที่อัลกอริทึมนี้สองทำได้ในชุดข้อมูลที่ต่างกัน

Batista และคณะ (2004) ได้ทำการศึกษาการจัดการกับชุดข้อมูลที่ไม่สมดุล โดยได้ทำการทดสอบโดยใช้ชุดข้อมูลจาก UCI Machine Learning Repository โดยใช้ Decision tree ในการวัดประสิทธิภาพ ซึ่งผลการทดลองได้ข้อสรุปว่าวิธีการ over-sampling ให้ผลดีกว่าวิธีการ under-sampling

Eichelberger และ Sheng (2013) ได้ทำการทดสอบวิเคราะห์เทคนิคการจำแนกหลายแบบ คลาสในชุดข้อมูลจาก UCI Machine Learning Repository โดยมีเปรียบเทียบประสิทธิภาพของอัลกอริทึม Decision Trees และ Naive Bayes ซึ่งผลการทดสอบคือ Decision Trees มีความแม่นยำสูงที่สุด และใช้เวลาในการฝึกสอนน้อยที่สุด แต่ขนาดของต้นไม้จะใหญ่เมื่อมีข้อมูลและคลาสมากขึ้น

Fauzi (2018) ได้ทำการศึกษาโดยใช้ Random Forest เพื่อจำแนกความรู้สึกที่อยู่ในภาษา จากข้อมูล 386 บทวิจารณ์ ซึ่งใช้คุณลักษณะของคำ ร่วมกับวิธีการให้น้ำหนักคำต่างๆ เช่น Binary TF, Raw TF, Logarithmic TF และ TF-IDF ได้ข้อสรุปคือ Random Forest เป็นเทคนิคที่มีประสิทธิภาพสำหรับการวิเคราะห์ข้อความ และวิธีการให้น้ำหนักคำที่ต่างกัน ไม่มีผลกระทบมากนักต่อผลลัพธ์

Flores และคณะ (2018) ได้ทำการประเมินและเปรียบเทียบประสิทธิภาพของ SVM และ Naive Bayes บนชุดข้อมูลการวิเคราะห์ความรู้สึกจาก Twitter และ Facebook โดยมุ่งเน้นการใช้เทคนิค SMOTE ในการจัดการกับข้อมูลที่มีคลาสจำนวนน้อย ซึ่งผลการทดลองแสดงว่า SMOTE เป็นเทคนิคสำคัญที่มีผลให้ความแม่นยำสูงขึ้นสำหรับ SVM และ Naive Bayes จากการแบ่งข้อมูลเพื่อฝึกและทดสอบที่อัตราส่วน 70 ต่อ 30

Raza และคณะ (2019) ได้วิเคราะห์เปรียบเทียบผลการทำงานของตัวแบบการเรียนรู้ของเครื่องสำหรับการจำแนกข้อความหลายคลาสจากข้อมูลรีวิวของผู้ใช้บริการ SaaS โดยอ้างอิงถึงเกณฑ์มาตรฐานของคุณภาพบริการ SaaS ที่กำหนดโดย AWS และ Microsoft Azure ใช้เทคนิค SMOTE เพื่อเพิ่มจำนวนตัวอย่างเทียมของคลาสที่มีจำนวนน้อย เพื่อลดปัญหาความไม่สมดุลของข้อมูล และวัดประสิทธิภาพด้วยเมตริกต่างๆ เช่น accuracy, precision, recall, F1-score เป็นต้น

K. A. Qureshi และ M. Sabih (2021) ได้ศึกษาเกี่ยวกับการทำตัวแบบการจำแนกข้อความแบบหลายคลาสประเภทคำพูดเกลียดชังในสื่อสังคมออนไลน์ โดยรวบรวมข้อมูลจากทวิตเตอร์และแบ่งเนื้อหาออกเป็น 10 คลาสของคำพูดเกลียดชัง จากการประเมินจากผู้เชี่ยวชาญจำนวน 12 คน ซึ่งได้ทำการเปรียบเทียบผลลัพธ์ของอัลกอริทึมต่างๆ เช่น Logistic Regression, Multilayer Perceptron, Decision Tree, SVM, Random Forest, CATBoost เป็นต้น ซึ่งผลการทดลองแสดงว่า แบบจำลอง

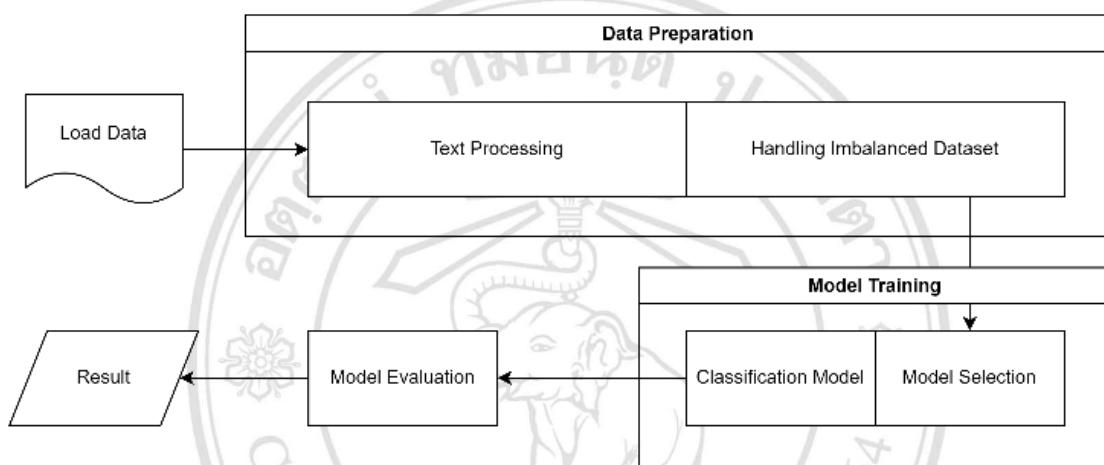
ลิขสิทธิ์ของมหาวิทยาลัยมหาสารคาม
CATBoost ให้ผลลัพธ์ที่ดีที่สุดกับข้อมูลชุดนี้

ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

บทที่ 3

วิธีการดำเนินการ

เทคนิคการจำแนกข้อความแบบหลายคลาส สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ มีกระบวนการทดสอบดังภาพที่ 3.1



ภาพที่ 3.1 : กระบวนการทดสอบ

โดยภาพรวมของกระบวนการทดสอบจะเป็นการนำชุดข้อมูลของศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ มาทำการเตรียมข้อมูลก่อนนำไปฝึกตัวแบบ ซึ่งต้องผ่านขั้นตอนการประมวลผลข้อความเพื่อตัดคำที่ไม่จำเป็นออกจากขั้นตอนการประมวลผล เนื่องจากชุดข้อมูลในแต่ละคลาสมีจำนวนไม่เท่ากัน จึงต้องผ่านการจัดการชุดข้อมูลที่ไม่สมดุลเพื่อลดอัตราของการประมวลผล เมื่อถึงขั้นตอนการฝึกตัวแบบ จะเป็นการนำข้อมูลประมวลผลเข้าทดสอบผ่านเทคนิคการจำแนกประเภทข้อความชนิดต่างๆ ก่อนแบ่งข้อมูลออกเป็น 2 ส่วนคือ ส่วนแรกเป็นข้อมูลที่ถูกจำแนกเป็นคลาสต่างๆ ไว้แล้วเพื่อฝึกให้ตัวแบบจำแนกข้อความได้ผ่านการเก็บข้อมูลที่มีอยู่ในแต่ละคลาส และส่วนที่สองเป็นข้อมูลที่ยังไม่ระบุคลาสที่จะถูกนำไปประนูคลาส จากนั้นจะทำการเปรียบเทียบข้อมูลทั้งสองส่วนว่ามีคลาสตรงกันเพียงใด เพื่อประเมินประสิทธิภาพของการจำแนกข้อความของตัวแบบนี้ โดยมีรายละเอียดของแต่ละกระบวนการดังต่อไปนี้

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ สงวนลิขสิทธิ์

ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

3.1 การเตรียมข้อมูล

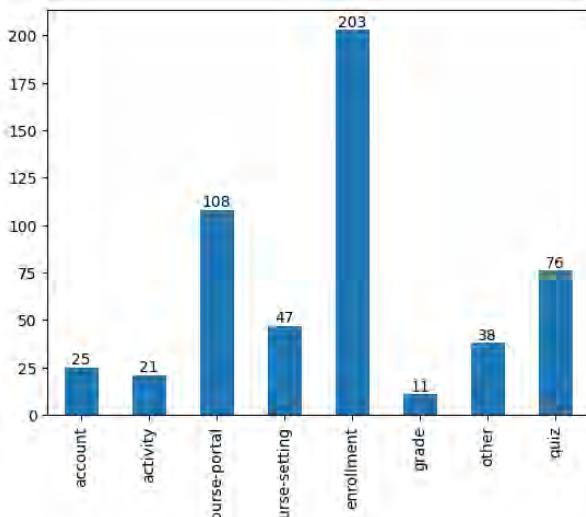
การเตรียมข้อมูล (Data Preparation) มีขั้นตอนดังนี้

3.1.1 การประมวลผลข้อความ (Text Processing) มีกระบวนการดังตารางที่ 3.1 ตารางที่ 3.1 การประมวลผลข้อความ (Text Processing)

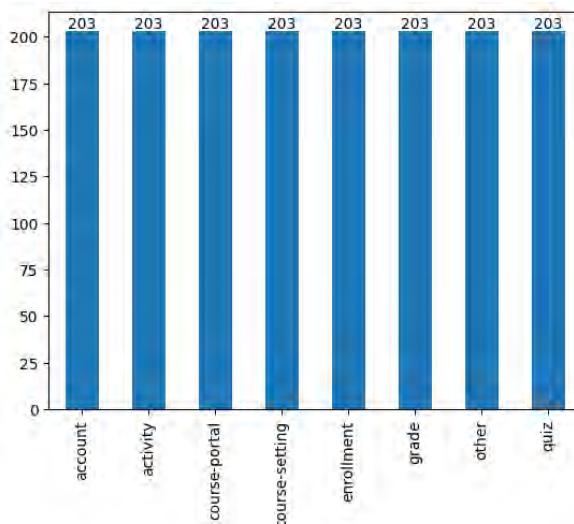
กระบวนการ	การดำเนินการ
ตัด format เวลา	ตัดส่วนระบุวันและเวลาในข้อมูล
ตัด email	ตัดอีเมลออกจากข้อความ
ตัด HTML + markup Code	ในการส่งอีเมลหรือบางข้อความมักจะมี HTML หรือ Markup Language ติดมาด้วย จึงตัดออกเพื่อให้คำชี้แจงกันมากขึ้น
ตัดข้อความที่ไม่ต้องการ	ลดช่องว่างที่เกิดขึ้น โดยเฉพาะหลังจากการตัด Code ออกไป
ตัดตัวเลข	มักจะเป็นตัวเลขที่เหลืออยู่หลังจากการตัด Code
ตัดคำ	โดยใช้ PyThaiNLP
ตัดชื่อคน	นำชื่อคนออกจากข้อความ
ตัดคำเนิพะ	ตัดคำเนิพะออกจากข้อความ เช่น ชื่องค์กร หรือชื่อคณะต่างๆ
ตัดคำฟุ่มเฟือย	ตัดคำฟุ่มเฟือยกออก

โดยข้อความหลังการประมวลผล จะถูกนำมาทดสอบการประเมินความสำคัญของคำใน
ข้อความของแต่ละคลาส โดยในการทดลองนี้มีการนำวิธีเวกเตอร์ของนับคำ (Count
ดาวน์โหลดเมื่อ 27/11/2567 09:00:00 AM) คำนวณโดยใช้ Vectorization (CV)⁵ และวิธีการหาค่าน้ำหนักของคำในเอกสาร (Term Frequency-Inverted Document Frequency : TF-IDF) มาทำการเปรียบเทียบกัน

3.1.2 จัดการกับชุดข้อมูลที่ไม่สมดุล (Handling Imbalanced Dataset) หลังจากประเมินความสำคัญของข้อความแต่ละคลาส จะมีการทดสอบเพื่อเปรียบเทียบเทคนิคการสังเคราะห์ข้อมูลเพิ่มเติม (Synthetic Minority Over-sampling Technique: SMOTE) และ Bootstrapping ในการจัดการชุดข้อความที่ไม่สมดุล โดยมีตัวอย่างชุดข้อมูลที่ข้อความทดสอบในแต่ละคลาสไม่เท่ากันดังภาพที่ 3.2 โดยมีคลาส enrollment เป็นคลาสที่มีจำนวนข้อความมากที่สุด และคลาส grade เป็นคลาสที่มีจำนวนข้อความน้อยที่สุด และหลังจากใช้เทคนิค SMOTE ที่มีจำนวนเท่ากันในแต่ละคลาส สังเคราะห์ข้อความขึ้นมาเพิ่มเติมในแต่ละคลาส จัดการกับชุดข้อมูลที่ไม่สมดุลโดยกำหนด random_state=42 เพื่อควบคุมความแปรปรวนของผลลัพธ์ตัวแบบ ให้ตัวแบบสร้างผลลัพธ์เดียวกัน เมื่อทำการทดสอบหลายครั้ง ซึ่งจะได้ผลลัพธ์ ดังภาพที่ 3.3



ภาพที่ 3.2 : แสดงตัวอย่างจำนวนข้อความที่ไม่เท่ากันในแต่ละคลาส
Copyright © by Chiang Mai University
All rights reserved



ภาพที่ 3.3 : แสดงตัวอย่างจำนวนข้อความหลังจากใช้เทคนิค SMOTE

3.2 การฝึกตัวแบบ

การฝึกตัวแบบ (Model Training) มีขั้นตอนดังนี้

3.2.1 การเลือกตัวแบบ (Model Selection) มีการทดสอบจากอัลกอริทึมการจำแนกข้อความ ได้แก่ Naïve Bayes, Logistic Regression, Support Vector Machine, Random Forest, CatBoost, Multilayer Perceptron, Decision Tree ซึ่งได้ทำการระบุพารามิเตอร์ดังตารางที่ 3.2

ตารางที่ 3.2 อัลกอริทึมและพารามิเตอร์

อัลกอริทึม	พารามิเตอร์
Naïve Bayes	MultinomialNB()
Logistic Regression	LogisticRegression(max_iter=10000)
Support Vector Machine	SVC()
Random Forest	RandomForestClassifier()
CatBoost	CatBoostClassifier(verbose=0,task_type="GPU")
Multilayer Perceptron	MLPClassifier(max_iter=1000)

ตารางที่ 3.2 อัลกอริทึมและพารามิเตอร์ (ต่อ)

อัลกอริทึม	พารามิเตอร์
Decision Tree	DecisionTreeClassifier()

3.2.2 การจำแนกข้อความ (Classification) เป็นขั้นตอนการฝึกตัวแบบที่มีการแบ่งชุดข้อมูลออกเป็น 2 ส่วนเพื่อทำการฝึก (Training) และทำการทดสอบ (Testing) ในอัตราส่วน 70:30 เพื่อใช้กับอัลกอริทึมการจำแนกข้อความที่เลือกไว้ก่อนหน้า

3.3 การประเมินตัวแบบ

การประเมินตัวแบบ (Model Evaluation) มีรายละเอียดดังต่อไปนี้

หลังจากทำการทดสอบแต่ละอัลกอริทึมทั้ง 7 ในขั้นตอนนี้จะเป็นการนำผลลัพธ์ที่ได้มาเปรียบเทียบประสิทธิภาพของตัวแบบโดยใช้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าความครบถ้วน (Recall) และ คะแนน F1 (F1-score) เป็นตัวชี้วัดประสิทธิภาพของแต่ละตัวแบบ และสรุปผลตัวแบบที่มีประสิทธิภาพสูงสุดจากค่าเฉลี่ยของการทดสอบทั้งหมด 5 รอบ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

บทที่ 4

ผลการทดลองและวิจารณ์ผลการทดลอง

จากการทำการทดลองด้วยการเขียนโปรแกรมภาษาไพธอนด้วย Colab Notebook โดยใช้จำนวนข้อมูลในคลาสต่างๆ ผลการทดสอบของการพัฒนาตัวแบบการจำแนกข้อความแบบหดายคลาสด้วยชุดข้อมูลที่ไม่สมดุลในการพัฒนาระบบบริการตนเอง สำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ว่าจะอัลกอริทึม Naïve Bayes, Logistic Regression, Support Vector Machine, Random Forest, CatBoost, Multilayer Perceptron, Decision Tree ทั้งหมด 5 รอบได้ค่าเฉลี่ยของผลการทดลองดังนี้

4.1 ผลการทดสอบความแม่นยำ

ผลการทดสอบความแม่นยำ (Accuracy) อัลกอริทึม Multilayer Perceptron ด้วยชุดข้อมูลสมดุลด้วยเทคนิค TF-IDF กับ SMOTE ให้ค่าความแม่นยำเฉลี่ย 95% ทำให้ Multilayer Perceptron เป็นอัลกอริทึมที่มีความแม่นยำมากที่สุดในกลุ่มทดสอบนี้ ดังค่าผลลัพธ์ในตารางที่ 4.1

ตารางที่ 4.1 : แสดงผลการทดสอบความแม่นยำ

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Naïve Bayes	0.53	0.77	0.66	0.38	0.61	0.90
Logistic Regression	0.55	0.78	0.72	0.43	0.72	0.94
Support Vector Machine	0.43	0.72	0.61	0.40	0.71	0.93

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ตารางที่ 4.1 : แสดงผลการทดสอบความแม่นยำ (ต่อ)

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Random Forest	0.48	0.75	0.71	0.47	0.75	0.92
CatBoost	0.52	0.75	0.69	0.53	0.74	0.90
Multilayer Perceptron	0.51	0.77	0.71	0.52	0.76	0.95
Decision Tree	0.43	0.71	0.65	0.40	0.70	0.81

4.2 ผลการทดสอบความเที่ยงตรง

ผลการทดสอบความเที่ยงตรง (Precision) อัลกอริทึม Support Vector Machine และ Multilayer Perceptron คำนวณข้อมูลสมดุลด้วยเทคนิค TF-IDF กับ SMOTE ให้ค่าความเที่ยงตรงเฉลี่ย 96% ทำให้ Support Vector Machine และ Multilayer Perceptron เป็นอัลกอริทึมที่มีความเที่ยงตรงมากที่สุดร่วมกันในกลุ่มทดสอบนี้ ดังค่าผลลัพธ์ในตารางที่ 4.2

ตารางที่ 4.2 : แสดงผลการทดสอบความเที่ยงตรง

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Naïve Bayes	0.39	0.49	0.71	0.15	0.26	0.91
Logistic Regression	0.43	0.55	0.75	0.28	0.33	0.94

ตารางที่ 4.2 : แสดงผลการทดสอบความเที่ยงตรง (ต่อ)

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Support Vector Machine	0.23	0.33	0.70	0.34	0.44	0.96
Random Forest	0.46	0.51	0.74	0.40	0.50	0.93
CatBoost	0.42	0.56	0.72	0.48	0.60	0.91
Multilayer Perceptron	0.41	0.54	0.75	0.44	0.59	0.96
Decision Tree	0.31	0.46	0.67	0.27	0.39	0.82

4.3 ผลการทดสอบค่าความครบถ้วน

ผลการทดสอบค่าความครบถ้วน (Recall) อัลกอริทึม Multilayer Perceptron ด้วยชุดข้อมูล สมดุลด้วยเทคนิค TF-IDF กับ SMOTE ให้ค่าความครบถ้วน 96% ทำให้ Multilayer Perceptron เป็น อัลกอริทึมที่มีความครบถ้วนมากที่สุดในกลุ่มทดสอบนี้ ดังค่าผลลัพธ์ในตารางที่ 4.3

ตารางที่ 4.3 : แสดงผลการทดสอบค่าความครบถ้วน

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Naïve Bayes	0.30	0.41	0.67	0.15	0.21	0.91

ตารางที่ 4.3 : แสดงผลการทดสอบค่าความครบถ้วน (ต่อ)

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Logistic Regression	0.35	0.46	0.72	0.18	0.30	0.94
Support Vector Machine	0.17	0.30	0.62	0.16	0.29	0.94
Random Forest	0.24	0.35	0.71	0.23	0.35	0.92
CatBoost	0.29	0.41	0.69	0.34	0.41	0.91
Multilayer Perceptron	0.31	0.45	0.71	0.32	0.44	0.96
Decision Tree	0.29	0.39	0.65	0.27	0.39	0.82

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
 Copyright© by Chiang Mai University
 All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
 ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

4.4 ผลการทดสอบคะแนน F1

ผลการทดสอบคะแนน F1 (F1-score) อัลกอริทึม Multilayer Perceptron ด้วยชุดข้อมูลสมดุล ด้วยเทคนิค TF-IDF กับ SMOTE ให้ค่าคะแนน F1 95% ทำให้ Multilayer Perceptron เป็นอัลกอริทึม ที่มีคะแนน F1 มากที่สุดในกลุ่มทดสอบนี้ ดังค่าผลลัพธ์ในตารางที่ 4.4

ตารางที่ 4.4 : แสดงผลการทดสอบคะแนน F1

อัลกอริทึม	CV	CV/Boot strapping	CV /SMOTE	TF-IDF	TF-IDF /Boot strapping	TF-IDF /SMOTE
Naïve Bayes	0.30	0.41	0.67	0.10	0.19	0.90
Logistic Regression	0.35	0.47	0.72	0.16	0.28	0.94
Support Vector Machine	0.14	0.27	0.63	0.13	0.27	0.94
Random Forest	0.25	0.36	0.72	0.23	0.35	0.92
CatBoost	0.31	0.42	0.69	0.35	0.43	0.91
Multilayer Perceptron	0.32	0.46	0.71	0.33	0.46	0.95
Decision Tree	0.28	0.39	0.65	0.25	0.38	0.82

4.5 วิจารณ์ผลการทดสอบ

จากการศึกษานี้ผู้ค้นคว้าได้ทดสอบตัวแบบเพื่อจำแนกข้อความแบบหลายคลาสคั่วychucxomud ที่ไม่สมดุล โดยแยกการทดสอบในขั้นตอนประเมินความสำคัญของคำในข้อความของแต่ละคลาส และจัดการกับชุดข้อมูลที่ไม่สมดุล ออกเป็น 4 การทดสอบคือ การทดสอบความแม่นยำ การทดสอบความเที่ยงตรง การทดสอบความครบถ้วน และการทดสอบคะแนน F1 โดยจากการทดสอบ 5 ครั้ง ผลการทดสอบเฉลี่ยแสดงให้เห็นว่าการใช้เทคนิค TF-IDF ร่วมกับ SMOTE ทำให้ประสิทธิภาพในการรวมของเทคนิคทั้งหมดที่ใช้ทดสอบมีผลลัพธ์ที่สูงขึ้น และอัลกอริทึม Multilayer Perceptron มีประสิทธิภาพในการจำแนกข้อความได้ดีที่สุด โดยค่าความแม่นยำ ค่าความเที่ยงตรง ค่าความครบถ้วน และคะแนน F1 สูงสุดทดสอบได้มีค่าอยู่ที่ 96% 96% 96% และ 95% ตามลำดับ สอดคล้องกับงานวิจัยของ Revina และคณะ (2020) ที่ผลของใช้ TF-IDF กับ วิธีการการเรียนรู้ของเครื่องมีประสิทธิภาพดีที่สุด โดยใช้เกณฑ์การวัดผลข้างต้น



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

บทที่ 5

สรุปผลการทดลอง

จากผลการทดลองทำให้เห็นว่ากรณีของชุดข้อมูลของศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ การสังเคราะห์ข้อมูลเพิ่มเติมมีโอกาสทำให้กลุ่มข้อมูลที่ถูกนับจากขั้นตอนประเมินความสำคัญของคำในข้อความของแต่ละคลาส โดย Multilayer Perceptron ที่ใช้ TF-IDF กับ SMOTE มีประสิทธิภาพในการจำแนกข้อความ ได้ดีที่สุด ซึ่งเป็นไปในทิศทางเดียวกันกับผลการวิจัยของ Flores และคณะ (2018) ที่อัลกอริทึมการเรียนรู้ของเครื่องได้รับการปรับปรุงประสิทธิภาพหลังจากที่ใช้ SMOTE ใน การประมวลผลข้อความ ทำให้มีความเหมาะสมในการใช้จำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่ไม่สมดุลในการพัฒนาระบบบริการติดต่องานสำหรับศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่

5.1 ข้อเสนอแนะ

5.1.1 ข้อเสนอแนะในการนำผลวิจัยไปใช้

- 1) นำตัวแบบไปพัฒนาระบบสนับสนุนผู้ใช้แบบบูรณาการเพลตฟอร์มของศูนย์นวัตกรรมการสอนและการเรียนรู้ ของมหาวิทยาลัยเชียงใหม่ ในส่วนของการวิเคราะห์ปัญหา เพื่อเข้าใจปัญหาในเชิงลึก จากการหาแนวโน้มและคาดการณ์ ช่วงเวลาการเกิดปัญหา และในส่วนของการแนะนำวิธีการช่วยเหลืออัตโนมัติ เพื่อเพิ่มประสิทธิภาพของการสนับสนุนการบริการด้วยตัวเอง เพื่อช่องทางการติดต่อกับผู้ใช้ ลดข้อจำกัดด้านเวลาทำการให้กับเจ้าหน้าที่ผู้ให้การสนับสนุน
- 2) กรณีนำตัวแบบไปใช้กับงานอื่นๆ ซึ่งบริบทของการใช้ตัวแบบนี้สามารถนำไปใช้ กับระบบที่มีลักษณะแบบเดียวกับข้อ 1) ในหน่วยงานอื่นๆ แต่สิ่งที่ต้องคำนึงถึง คือเวลาที่ใช้ประมวลผลจะเปลี่ยนไปตามจำนวนข้อความ ประเภทของตัวแบบที่จะถูกเปลี่ยนไปตามคุณลักษณะของข้อมูลที่หน่วยงานได้รับ และสมรรถนะของ เครื่องคอมพิวเตอร์ประมวลผล นอกจากนี้สามารถนำไปใช้ในงานด้านการค้นหาโดยใช้เครื่องค้นหาน้ำหนัก 06/05/2568 โดย นางสาวอรุณรัตน์ ใจดี วันที่ 27/11/2567 09:39:17 ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ ได้รับ และสมรรถนะของ วิเคราะห์อารมณ์หรือความรู้สึกสำหรับงานที่มีการเก็บข้อมูลของลูกค้า หรือข้อมูลจากช่องทาง Social Media เป็นต้น

5.1.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

- 1) การทดลองกับชุดข้อมูลที่มีคลาสที่หลากหลายขึ้น ชุดข้อมูลที่มีการระบุคลาสที่มากขึ้นจะช่วยให้ประสิทธิภาพของการจำแนกข้อความมีมากขึ้น และสามารถจำแนกคลาสเพิ่มเติมจากคลาสเดิมได้ เช่น ระบุคลาส “ลีมรหัสผ่าน” “การเข้าสู่ระบบ” “ข้อมูลบัญชี” ได้เพิ่มเติมจากคลาสของปัญหาเกี่ยวกับ “บัญชีผู้ใช้” เพื่อการจำแนกข้อความที่แม่นยำมากขึ้น
- 2) การพัฒนาตัวแบบที่สามารถจำแนกข้อความ ได้หลายคลาสใน 1 ข้อความ
- 3) ผลการทดสอบเกิดจากการจำแนกข้อความแบบหลายคลาสด้วยชุดข้อมูลที่เกี่ยวข้องกับบริการของศูนย์นวัตกรรมการสอนและการเรียนรู้ มหาวิทยาลัยเชียงใหม่ หากนำไปใช้กับข้อมูลชุดอื่นควรปรับขั้นตอนการประมวลผลข้อความ และตัวแบบการจำแนกให้เหมาะสม

5.2 ปัญหาในการวิจัย

- 5.2.1 การทดสอบตัวแบบใช้เวลานานหาก GPU ถูกจำกัดการใช้งาน โดยเฉพาะการทดสอบของ CATBoost และ Multilayer Perceptron
- 5.2.2 ผลลัพธ์ของการทดสอบมีการเปลี่ยนแปลง หลังจากการอัปเดตเวอร์ชันของ Library การตัดคำ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

ເອກສານອ້າງອີງ (References)

- [Alexander Genkin et al., 2007] Alexander Genkin, Genkin, A., David D. Lewis, Lewis, D., Lewis, D. D., & Madigan, D. (2007). Large-Scale Bayesian Logistic Regression for Text Categorization. *Technometrics*, 49(3), 291–304. <https://doi.org/10.1198/004017007000000245>
- [Batista et al., 2004] Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *Sigkdd Explorations*, 6(1), 20–29. <https://doi.org/10.1145/1007730.1007735>
- [Chawla et al., 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1), 321–357. <https://doi.org/10.1613/jair.953>
- [Eichelberger & Sheng, 2013] Eichelberger, R. K., & Sheng, V. S. (2013). An empirical study of reducing multiclass classification methodologies. *IAPR International Conference on Machine Learning and Data Mining in Pattern Recognition*, 505–519. https://doi.org/10.1007/978-3-642-39712-7_39
- [Fauzi, 2018] Fauzi, M. A. (2018). Random Forest Approach for Sentiment Analysis in Indonesian Language. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1), 46–50. <https://doi.org/10.11591/ijeecs.v12.i1.pp46-50>

- [Flores et al., 2018] Flores, A. C., Icoy, R. I., Pena, C. F., & Gorro, K. D. (2018). An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set. *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, 1–4. <https://doi.org/10.1109/iceast.2018.8434401>
- [Frank & Bouckaert, 2006] Frank, E., & Bouckaert, R. R. (2006). Naive Bayes for text classification with unbalanced classes. *European Conference on Principles of Data Mining and Knowledge Discovery*, 503–510. https://doi.org/10.1007/11871637_49
- [Geoffrey Holmes et al., 2002] Geoffrey Holmes, Holmes, G., Geoffrey Holmes, Pfahringer, B., Kirkby, R., Frank, E., Mark A. Hall, Hall, M. A., & Hall, M. A. (2002). Multiclass alternating decision trees. *Lecture Notes in Computer Science*, 161–172. https://doi.org/10.1007/3-540-36755-1_14
- [Guan et al., 2019] Guan, M., Cho, S., Petro, R. M., Zhang, W., Zhang, W., Pasche, B., & Topaloglu, U. (2019). Natural language processing and recurrent network models for identifying genomic mutation-associated cancer treatment change from patient progress notes. *JAMIA Open*, 2(1), 139–149. <https://doi.org/10.1093/jamiaopen/ooy061>
- [Hancock et al., 2020] Hancock, J. T., Taghi M. Khoshgoftaar, Khoshgoftaar, T. M., Khoshgoftaar, T. M., & Taghi M. Khoshgoftaar. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 7(1), 1–45. <https://doi.org/10.1186/s40537-020-00369-8>

- [Joachims, 1998] Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *European Conference on Machine Learning*, 1398, 137–142. <https://doi.org/10.1007/bfb0026683>
- [K. A. Qureshi and M. Sabih, 2021] K. A. Qureshi and M. Sabih. (2021). Un-Compromised Credibility: Social Media Based Multi-Class Hate Speech Classification for Text. *IEEE Access*, vol. 9, pp. 109465–109477, 2021, <https://doi.org/10.1109/access.2021.3101977>
- [Kawamura et al., 2021] Kawamura, R., Shirai, S., Takemura, N., Alizadeh, M., Cukurova, M., Takemura, H., & Nagahara, H. (2021). Detecting Drowsy Learners at the Wheel of e-Learning Platforms With Multimodal Learning Analytics. *IEEE Access*, vol. 9, 115165–115174. <https://doi.org/10.1109/access.2021.3104805>
- [Kotsiantis, 2007] Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica (Lithuanian Academy of Sciences)*, 31(3), 249–268. <https://dl.acm.org/doi/10.5555/1566770.1566773>
- [Meuter et al., 2000] Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters. *Journal of Marketing*, 64(3), 50-64. <https://doi.org/10.1509/jmkg.64.3.50.18024>
- [Prokhorenkova et al., 2018] Prokhorenkova, L. O., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Neural Information Processing Systems*, 31, 6639–6649.

- [Raza et al., 2019] Raza, M., Hussain, F. K., Hussain, O. K., Zhao, M., Khan, I. A., & Rehman, Z. ur. (2019). A comparative analysis of machine learning models for quality pillar assessment of SaaS services by multi-class text classification of users' reviews. *Future Generation Computer Systems*, 101, 341–371. <https://doi.org/10.1016/j.future.2019.06.022>
- [Rennie et al., 2003] Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (2003). *Tackling the poor assumptions of naive bayes text classifiers*. 616–623.
- [Revina et al., 2020] Revina, A., Buza, K., Buza, K., & Meister, V. G. (2020). IT Ticket Classification: The Simpler, the Better. *IEEE Access*, 8, 193380–193395. <https://doi.org/10.1109/access.2020.3032840>
- [Rokach, 2016] Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, 27(27), 111–125. <https://doi.org/10.1016/j.inffus.2015.06.005>
- [Sánchez-Marcano et al., 2009] Sánchez-Marcano, N., Alonso-Betanzos, A., & Calvo-Estevez, R. M. (2009). *A Wrapper Method for Feature Selection in Multiple Classes Datasets*. 456–463. https://doi.org/10.1007/978-3-642-02478-8_57
- [Sun et al., 2007] Sun, Y., Kamel, M. S., Wong, A., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12), 3358–3378. <https://doi.org/10.1016/j.patcog.2007.04.009>
- [Vlachos, 2008] Vlachos, A. (2008). A stopping criterion for active learning. *Computer Speech & Language*, 22(3), 295–312. <https://doi.org/10.1016/j.csl.2007.12.001>

- [Xu, 2018] Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1), 48–59. <https://doi.org/10.1177/0165551516677946>
- [Yanjiao Xiong et al., 2022] Yanjiao Xiong, Zhipeng Zhao, & Zhigang Duan. (2022). *A study of Thai word separation methods based on machine learning*. <https://doi.org/10.21203/rs.3.rs-2064323/v1>



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ภาคผนวก ก

ตัวอย่างของกระบวนการเตรียมข้อมูล

ตารางที่ ก.1 : ตารางการเตรียมข้อมูล

กระบวนการ	วิธีที่ใช้	คำอธิบาย
ตัด format เวลา	ใช้ Regex " <code>\b(?:1[0-2] 0?[1-9]):[0-5][0-9](?:[0-5][0-9])?(?: [APap][mM])?\b'</code>	ตัด format ของเวลา และวันที่ออกจากข้อความ
ตัด email	ใช้ Regex " <code>\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Z az]{2,7}\b'</code>	ตัด email ออกจากข้อความโดยนำเงื่อนไขตัดอักษรที่อยู่ติดกับเครื่องหมาย @
ตัด HTML + markup Code	ใช้ Regex - ' <code>\{[^{}]*\}'</code> - ' <code>\[[^\]]*\]</code> ' - ' <code>\{[^{}]*\}</code> ' - ' <code>\[[^\]]*\]</code> ' - ' <code>https?:\/\/S+ www\.S+</code> '	ตัดอักษรที่อยู่ในเครื่องหมาย {}, [] และ URL
ตัดชื่นบรรทัดใหม่	ใช้ rm_useless_newline ของ PyThaiNLP	ทำให้ข้อความรวมอยู่ในบรรทัดเดียวกัน
ตัดตัวเลข	ใช้ Regex '[0-9]'	ตัดตัวเลข 0-9 ออกจากข้อความ
ตัดคำ	ใช้ tokenize ของ PyThaiNLP	ทำการแบ่งคำจากข้อความให้ออกมาเป็นคำหรืออวสาน

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย์
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ตารางที่ ก.1 : ตารางการเตรียมข้อมูล (ต่อ)

กระบวนการ	วิธีที่ใช้	คำอธิบาย
ตัดชื่อคน	ใช้ thai_family_names(), thai_female_names, thai_male_names() ของ PyThaiNLP	ตัดชื่อคนออกจากข้อความ
ตัดคำเฉพาะ	ใช้กลุ่มคำเฉพาะ เช่น ชื่องค์กร ชื่อคณะ ภาควิชา เป็นต้น	ตัดชื่อเฉพาะออกจากข้อความ
ตัดคำฟู่มเพือย	ใช้ thai_stopwords ของ PyThaiNLP และ eng_stopwords ของ nltk	ตัดคำฟู่มเพือยออกจาก ข้อความ

ตัวอย่างข้อความ

“การเพิ่มชื่อนักศึกษา เรียน admin มีนักศึกษาแจ้งว่าไม่มีชื่อใน [REDACTED] ทำให้ไม่สามารถเข้าไปทำแบบฝึกหัดหรืออบรมทดสอบได้ รบกวนเติมชื่อนักศึกษาให้ด้วยค่ะ * [REDACTED] {color:#242424} * [REDACTED] {color:#242424} ขอบคุณมากค่ะ {color:#242424} {color:#242424} [REDACTED]”

ผลลัพธ์

“การเพิ่มชื่อนักศึกษา เรียน admin มีนักศึกษาแจ้งว่าไม่มีชื่อใน [REDACTED] ทำให้ไม่สามารถเข้าไปทำแบบฝึกหัดหรืออบรมทดสอบได้ รบกวนเติมชื่อนักศึกษาให้ด้วยค่ะ * [REDACTED] {color:#242424} ขอบคุณมากค่ะ * [REDACTED] {color:#242424}”

ภาคผนวก ข

เปรียบเทียบระยะเวลาการประมวลผลโดยใช้หน่วยประมวลผลกราฟิก ของ CatBoost

เปรียบเทียบเวลาในการประมวลผลโดยใช้หน่วยประมวลผลกราฟิกของ CatBoost ประมวลผลโดยใช้เครื่องที่มีทรัพยากรต่อไปนี้

1. เครื่อง A – RAM 12.7 GB / Disk 107.7 GB
2. เครื่อง B – RAM 12.7 GB / GPU RAM 15.0 GB / Disk 78.2 GB

ตารางที่ ข.1 : ตารางระยะเวลาการประมวลผลโดยใช้หน่วยประมวลผลกราฟิก

Model	เครื่อง A (นาที)	เครื่อง B (นาที)	แตกต่าง(นาที)
CatBoost + CV	69.29	5.93	63.35
CatBoost + CV + Bootstrap	83.97	7.02	76.95
CatBoost + CV + SMOTE	70.16	5.87	64.29
CatBoost + TFIDF	86.62	7.10	79.51
CatBoost + CV + Bootstrap	71.72	6.31	65.41
CatBoost + SMOTE + TFIDF	73.29	12.89	60.40
รวม	455.05	45.12	409.93

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ภาคผนวก ก

เปรียบเทียบผลการทดสอบก่อน-หลังการอัปเดตไอลบรารี

ก่อนการอัปเดตไอลบรารี - pythainlp (3.1.1) / catboost (1.1.1)

พบว่าการใช้ SMOTE ร่วมกับ TF-IDF ด้วยโมเดล SVM ได้ผลลัพธ์ที่ดี ความแม่นยำ ความเที่ยงตรง ความครบถ้วน และคะแนน F1 ดีที่สุดกับข้อมูลชุดนี้ ดังตารางที่ ค.1

ตารางที่ ค.1 : ตารางผลการทดสอบก่อนอัปเดตไอลบรารี

อัลกอริทึม	ความแม่นยำ	ความเที่ยงตรง	ความครบถ้วน	คะแนน F1
Naïve Bayes	0.85	0.86	0.85	0.85
Logistic Regression	0.89	0.90	0.90	0.90
Support Vector Machine	0.94	0.94	0.94	0.94
Random Forest	0.90	0.91	0.91	0.91
CatBoost	0.88	0.89	0.90	0.89
Multilayer Perceptron	0.92	0.92	0.92	0.91
Decision Tree	0.82	0.82	0.82	0.81

หลังการอัปเดตไลบรารี - pythainlp (4.0.2) / catboost (1.2.2)

พบว่าการใช้ SMOTE ร่วมกับ TF-IDF ด้วยโมเดล Multilayer Perceptron ได้ผลลัพธ์ที่ความแม่นยำ
ความเที่ยงตรง ความครบถ้วน และคะแนน F1 ดีที่สุดกับข้อมูลชุดนี้ ดังตารางที่ ค.2

ตารางที่ ค.2 : ตารางผลการทดสอบหลังอัปเดตไลบรารี

อัลกอริทึม	ความแม่นยำ	ความเที่ยงตรง	ความครบถ้วน	คะแนน F1
Naïve Bayes	0.90	0.91	0.91	0.90
Logistic Regression	0.94	0.94	0.94	0.94
Support Vector Machine	0.93	0.96	0.94	0.94
Random Forest	0.92	0.93	0.92	0.92
CatBoost	0.90	0.91	0.91	0.91
Multilayer Perceptron	0.95	0.96	0.96	0.95
Decision Tree	0.81	0.82	0.82	0.82

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวชิรวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568

ประวัติผู้เขียน

ชื่อ-นามสกุล

นาย ปณิธ พันธุ์วุฒิ

ประวัติการศึกษา ปีการศึกษา 2552 วิทยาศาสตรบัณฑิต สาขาวิชาบริหารคอมพิวเตอร์
มหาวิทยาลัยเชียงใหม่

ประวัติการทำงาน นักวิชาการคอมพิวเตอร์ สำนักบริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยเชียงใหม่
นักวิชาการคอมพิวเตอร์ ศูนย์นวัตกรรมการสอนและการเรียนรู้ สำนักงาน
มหาวิทยาลัย มหาวิทยาลัยเชียงใหม่



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่ โดย นายวิริวิทย์ จันทร์เพย
ดาวน์โหลดเมื่อ 27/11/2567 09:39:17 และหมดอายุ 06/05/2568