

บทที่ 3

ขั้นตอน และ วิธีการดำเนินงาน

โครงการเรื่อง การเปรียบเทียบตัวแบบที่เหมาะสมสำหรับการพยากรณ์การใช้ น้ำประปาด้วยเทคนิคเหมืองข้อมูล และนำเสนอข้อมูลผ่านเว็บไซต์ ในบทนี้จะมีขั้นตอนและ รายละเอียดดังนี้ โดยทำการเปรียบเทียบตัวแบบ 4 ตัวแบบ คือเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network : ANN) เทคนิคการวิเคราะห์ข้อมูลถดถอยเชิงเส้น (Linear Regression) เทคนิคแรนดอมฟอเรสต์ (Random Forest) และเทคนิคต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees) และนำมาเปรียบเทียบประสิทธิภาพของตัวแบบ โดยใช้ค่า RMSE และ MAE เป็นตัวเปรียบเทียบ จากนั้นทำการนำตัวแบบที่ได้ไปทำการพัฒนาเว็บไซต์ สำหรับการแสดงผล

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

เอกสิทธิ์ พัทธวงศ์ศักดิ์ (2558) กล่าวว่ากระบวนการมาตรฐานสำหรับการวิเคราะห์ ข้อมูล Data Mining หรือที่เรียกว่า Cross-Industry Standard Process for Data Mining (CRISP-DM) ถูกพัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของสามบริษัท ได้แก่ DaimlerChrysler, SPSS และ NCR โดยกระบวนการนี้ประกอบด้วย 6 ขั้นตอนหลัก

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding)

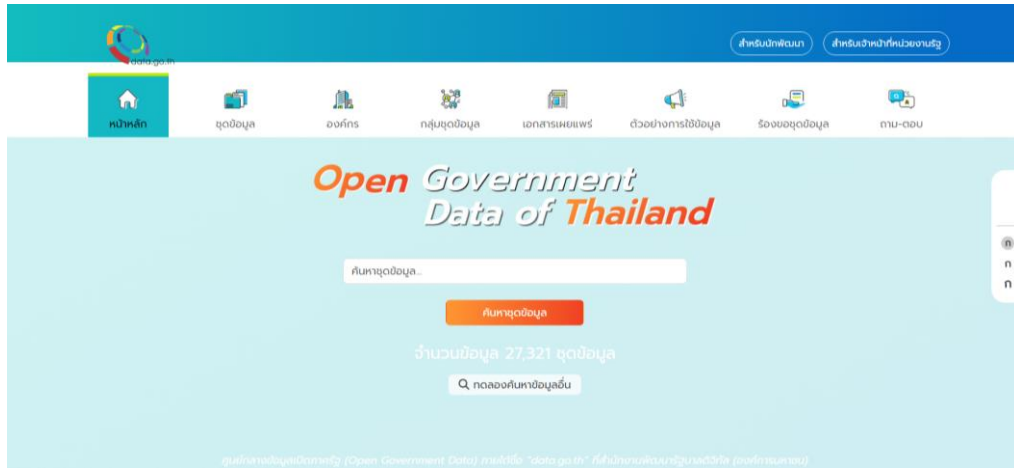
เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทาง ธุรกิจโดยรวมเป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการ ทางธุรกิจโดยรวม ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูล ทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ข้อมูลถูกจัดเก็บโดยศูนย์กลางข้อมูล เป็ดภาครัฐ และกรมประปานครหลวง

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้อง ของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับ วัตถุประสงค์ที่กำหนดไว้ผู้วิเคราะห์ทำการกำหนดไว้

ผู้ศึกษาทำการรวบรวมข้อมูล ข้อมูลปีพ.ศ. ข้อมูลเดือน ข้อมูลรหัสจังหวัด ข้อมูลชื่อ จังหวัด ข้อมูลรหัสอำเภอ ข้อมูลชื่ออำเภอ ข้อมูลรหัสตำบล ข้อมูลชื่อตำบล ข้อมูลรหัสประเภท

ผู้ใช้งาน ข้อมูลชื่อประเภทผู้ใช้งาน และปริมาณการใช้น้ำ เพื่อทำการวิเคราะห์ปริมาณการใช้น้ำประปา ทำการรวบรวมมาจากเว็บไซต์ของ ศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th) เข้าถึงได้จาก <https://data.go.th/> ซึ่งเป็นข้อมูลเปิดเผยต่อสาธารณะ



ภาพที่ 3.1 หน้าเว็บไซต์ data.go.th (ที่มา : <https://data.go.th/>)

ในไฟล์แต่ละปีจะมีการเก็บข้อมูลดังนี้ ข้อมูลปี ข้อมูลเดือน ข้อมูลรหัสจังหวัด ข้อมูลชื่อจังหวัด ข้อมูลรหัสอำเภอ ข้อมูลชื่ออำเภอ ข้อมูลรหัสตำบล ข้อมูลชื่อตำบล ข้อมูลรหัสประเภทผู้ใช้งาน ข้อมูลชื่อประเภทผู้ใช้งาน ปริมาณการใช้น้ำ ดังภาพที่ 3.2

PERIOD_YEAR	PERIOD_MONTH	PROVINCE_CODE	PROVINCE_NAME	AMPHUR_CODE	AMPHUR_NAME	DISTRICT_CODE	TAMBON_NAME	CLASS_GR CLASS_GROUP_NAME
2566	1	10	กรุงเทพมหานคร	1	พระนคร	1	พระบรมหาราชวัง	1 ที่อยู่อาศัย, ศาลากลาง, มุขนิย, นายของเล็กน้อย นายเนาที่อยู่อาศัย(R.1) ทั้งหมด
2566	1	10	กรุงเทพมหานคร	1	พระนคร	1	พระบรมหาราชวัง	2 ครัว, ร้านค้า, ธนาคาร, สถาบันการเงิน, สคิป, มีน่านิม, โรงพยาบาลเอกชน, นาย
2566	1	10	กรุงเทพมหานคร	1	พระนคร	1	พระบรมหาราชวัง	4 ที่อาคาร, สถาบันตั้ง, โรงเรียน, ศูนย์การค้า, ตลาด, สหกรณ์
2566	1	10	กรุงเทพมหานคร	1	พระนคร	1	พระบรมหาราชวัง	5 สถานที่ยาสูบ, องค์การระหว่างประเทศ, สถาบันของรัฐ, รัฐวิสาหกิจ, โรงพยาบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	1	พระบรมหาราชวัง	6 โรงเรียนเอกชน - รัฐบาล, วิทยาลัยเอกชน - รัฐบาล, มหาวิทยาลัยเอกชน - รัฐบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	1	พระบรมหาราชวัง	7 นำชั่วคราวก่อสร้าง
2566	1	10	กรุงเทพมหานคร	1	พระนคร	2	วังสุทโศทิม	1 ที่อยู่อาศัย, ศาลากลาง, มุขนิย, นายของเล็กน้อย นายเนาที่อยู่อาศัย(R.1) ทั้งหมด
2566	1	10	กรุงเทพมหานคร	1	พระนคร	2	วังสุทโศทิม	2 ครัว, ร้านค้า, ธนาคาร, สถาบันการเงิน, สคิป, มีน่านิม, โรงพยาบาลเอกชน, นาย
2566	1	10	กรุงเทพมหานคร	1	พระนคร	2	วังสุทโศทิม	4 ที่อาคาร, สถาบันตั้ง, โรงเรียน, ศูนย์การค้า, ตลาด, สหกรณ์
2566	1	10	กรุงเทพมหานคร	1	พระนคร	2	วังสุทโศทิม	5 สถานที่ยาสูบ, องค์การระหว่างประเทศ, สถาบันของรัฐ, รัฐวิสาหกิจ, โรงพยาบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	2	วังสุทโศทิม	6 โรงเรียนเอกชน - รัฐบาล, วิทยาลัยเอกชน - รัฐบาล, มหาวิทยาลัยเอกชน - รัฐบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	2	วังสุทโศทิม	7 นำชั่วคราวก่อสร้าง
2566	1	10	กรุงเทพมหานคร	1	พระนคร	3	วัดราชบพิธ	1 ที่อยู่อาศัย, ศาลากลาง, มุขนิย, นายของเล็กน้อย นายเนาที่อยู่อาศัย(R.1) ทั้งหมด
2566	1	10	กรุงเทพมหานคร	1	พระนคร	3	วัดราชบพิธ	2 ครัว, ร้านค้า, ธนาคาร, สถาบันการเงิน, สคิป, มีน่านิม, โรงพยาบาลเอกชน, นาย
2566	1	10	กรุงเทพมหานคร	1	พระนคร	3	วัดราชบพิธ	4 ที่อาคาร, สถาบันตั้ง, โรงเรียน, ศูนย์การค้า, ตลาด, สหกรณ์
2566	1	10	กรุงเทพมหานคร	1	พระนคร	3	วัดราชบพิธ	5 สถานที่ยาสูบ, องค์การระหว่างประเทศ, สถาบันของรัฐ, รัฐวิสาหกิจ, โรงพยาบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	3	วัดราชบพิธ	6 โรงเรียนเอกชน - รัฐบาล, วิทยาลัยเอกชน - รัฐบาล, มหาวิทยาลัยเอกชน - รัฐบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	3	วัดราชบพิธ	7 นำชั่วคราวก่อสร้าง
2566	1	10	กรุงเทพมหานคร	1	พระนคร	4	ศาลาญายอ	1 ที่อยู่อาศัย, ศาลากลาง, มุขนิย, นายของเล็กน้อย นายเนาที่อยู่อาศัย(R.1) ทั้งหมด
2566	1	10	กรุงเทพมหานคร	1	พระนคร	4	ศาลาญายอ	2 ครัว, ร้านค้า, ธนาคาร, สถาบันการเงิน, สคิป, มีน่านิม, โรงพยาบาลเอกชน, นาย
2566	1	10	กรุงเทพมหานคร	1	พระนคร	4	ศาลาญายอ	4 ที่อาคาร, สถาบันตั้ง, โรงเรียน, ศูนย์การค้า, ตลาด, สหกรณ์
2566	1	10	กรุงเทพมหานคร	1	พระนคร	4	ศาลาญายอ	5 สถานที่ยาสูบ, องค์การระหว่างประเทศ, สถาบันของรัฐ, รัฐวิสาหกิจ, โรงพยาบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	4	ศาลาญายอ	6 โรงเรียนเอกชน - รัฐบาล, วิทยาลัยเอกชน - รัฐบาล, มหาวิทยาลัยเอกชน - รัฐบาล
2566	1	10	กรุงเทพมหานคร	1	พระนคร	4	ศาลาญายอ	7 นำชั่วคราวก่อสร้าง
2566	1	10	กรุงเทพมหานคร	1	พระนคร	5	ศาลเจ้าพ่อเสือ	1 ที่อยู่อาศัย, ศาลากลาง, มุขนิย, นายของเล็กน้อย นายเนาที่อยู่อาศัย(R.1) ทั้งหมด
2566	1	10	กรุงเทพมหานคร	1	พระนคร	5	ศาลเจ้าพ่อเสือ	2 ครัว, ร้านค้า, ธนาคาร, สถาบันการเงิน, สคิป, มีน่านิม, โรงพยาบาลเอกชน, นาย
2566	1	10	กรุงเทพมหานคร	1	พระนคร	5	ศาลเจ้าพ่อเสือ	4 ที่อาคาร, สถาบันตั้ง, โรงเรียน, ศูนย์การค้า, ตลาด, สหกรณ์
2566	1	10	กรุงเทพมหานคร	1	พระนคร	5	ศาลเจ้าพ่อเสือ	5 สถานที่ยาสูบ, องค์การระหว่างประเทศ, สถาบันของรัฐ, รัฐวิสาหกิจ, โรงพยาบาล

ภาพที่ 3.2 ข้อมูลการใช้น้ำประปาที่ถูกเก็บบันทึกไว้

3.1.3 การเตรียมข้อมูล (Data Preparation)

3.1.3.1 การคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล ข้อมูลปริมาณการใช้น้ำในช่วงปีพ.ศ. 2557-พ.ศ.2566 โดยทำการเลือกข้อมูลที่จำเป็นในการวิเคราะห์ภาพรวมจากศูนย์กลางข้อมูลเปิดภาครัฐ จำนวน 6 แอดทริบิวท์ ได้แก่ ข้อมูลปีพ.ศ. ข้อมูลเดือน ข้อมูลรหัสจังหวัด ข้อมูลรหัสอำเภอ ข้อมูลรหัสประเภทผู้ใช้งาน และข้อมูลปริมาณการใช้น้ำประปา

3.1.3.2 การกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูล เป็น กระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

1) ข้อมูลปริมาณการใช้น้ำในช่วงปีพ.ศ. 2557-พ.ศ.2566 จากการวิเคราะห์พบว่า ข้อมูลที่ได้เป็นภาษาไทย ผู้ศึกษาจึงได้ทำการตรวจสอบข้อมูลและแก้ไขให้เป็นภาษาอังกฤษทั้งหมด

3.1.4 การสร้างตัวแบบ (Modeling)

3.1.4.1 การสร้างตัวแบบการทำนายโดยใช้เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network : ANN)

เทคนิค Artificial Neural Network จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

- 1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner
- 2) เลือกข้อมูลที่จะนำไปพยากรณ์ โดยใช้ operators ที่ชื่อว่า Select Attributes
- 3) ทำการกำหนดตัวแปรโดยใช้ operators ที่ชื่อว่า Set Role
- 4) สร้างตัวแบบ Artificial Neural Network โดยใช้ operators ที่ชื่อว่า Artificial Neural Network
- 5) แสดงผลการวัดประสิทธิภาพโดยใช้ operators ที่ชื่อว่า Performance (Regression)

3.1.4.2 การสร้างตัวแบบการทำนายโดยใช้เทคนิคการวิเคราะห์ข้อมูลถดถอยเชิงเส้น (Linear Regression)

เทคนิค Linear Regression จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

- 1) เลือกชุดข้อมูลที่นำเข้าสู่โปรแกรม Rapid Miner
- 2) เลือกข้อมูลที่จะนำไปพยากรณ์ โดยใช้ operators ที่ชื่อว่า Select Attributes

- 3) ทำการกำหนดตัวแปรโดยใช้ operators ที่ชื่อว่า Set Role
- 4) สร้างตัวแบบ Linear Regression โดยใช้ operators ที่ชื่อว่า Linear Regression
- 5) แสดงผลการวัดประสิทธิภาพโดยใช้ operators ที่ชื่อว่า Performance (Regression)

3.1.4.3 การสร้างตัวแบบการทำนายโดยใช้เทคนิคแรนดอมฟอเรสต์ (Random Forest)

เทคนิค Random Forest จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

- 1) เลือกชุดข้อมูลที่น่าเข้าสู่โปรแกรม Rapid Miner
- 2) เลือกข้อมูลที่จะนำไปพยากรณ์ โดยใช้ operators ที่ชื่อว่า Select Attributes
- 3) ทำการกำหนดตัวแปรโดยใช้ operators ที่ชื่อว่า Set Role
- 4) สร้างตัวแบบ Random Forest โดยใช้ operators ที่ชื่อว่า Random Forest
- 5) แสดงผลการวัดประสิทธิภาพโดยใช้ operators ที่ชื่อว่า Performance (Regression)

3.1.4.4 การสร้างตัวแบบการทำนายโดยใช้เทคนิคต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees)

เทคนิค Gradient Boosted Trees จะทำการสร้างและวัดผลในโปรแกรม RapidMiner

- 1) เลือกชุดข้อมูลที่น่าเข้าสู่โปรแกรม Rapid Miner
- 2) เลือกข้อมูลที่จะนำไปพยากรณ์ โดยใช้ operators ที่ชื่อว่า Select Attributes
- 3) ทำการกำหนดตัวแปรโดยใช้ operators ที่ชื่อว่า Set Role
- 4) สร้างตัวแบบ Gradient Boosted Trees โดยใช้ operators ที่ชื่อว่า Gradient Boosted Trees
- 5) แสดงผลการวัดประสิทธิภาพโดยใช้ operators ที่ชื่อว่า Performance (Regression)

3.1.5 การวัดประสิทธิภาพของตัวแบบ (Evaluation)

วัดผลประสิทธิภาพทั้ง 4 เทคนิค คือ เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network : ANN) เทคนิคการวิเคราะห์ข้อมูลถดถอยเชิงเส้นหรือ (Linear Regression) เทคนิคแรนดอมฟอเรสต์ (Random Forest) และเทคนิคต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees) โดยคำนวณค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) และ ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Absolute Error : MAE) และเลือกตัวแบบที่มีประสิทธิภาพดีที่สุดในใช้งานจริง

ตารางที่ 3.1 ตารางเปรียบเทียบประสิทธิภาพระหว่างตัวแบบ

ตัวแบบ	RMSE	MAE
Artificial Neural Network		
Linear Regression		
Random Forest		
Gradient Boosted Trees		

3.1.6 การนำตัวแบบไปใช้งานจริง (Deployment) ในการนำโมเดลที่ผ่านการพัฒนาและเปรียบเทียบมาใช้งานจริง มีการดำเนินการดังต่อไปนี้

1) การเลือกตัวแบบที่เหมาะสม

จากการวิเคราะห์ตัวแบบด้วยโปรแกรม RapidMiner ได้ทำการเปรียบเทียบค่าความคลาดเคลื่อน (Error) ของตัวแบบทั้ง 4 ตัวแบบ และเลือกตัวแบบที่มีค่าความคลาดเคลื่อนต่ำที่สุดโดยใช้เกณฑ์ Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)

2) การพยากรณ์และประยุกต์ใช้งาน

ตัวแบบที่ได้รับการเลือกจะถูกนำมาใช้ในการพยากรณ์ข้อมูลในอนาคต โดยอาศัยข้อมูลจากชุดข้อมูลใหม่ เพื่อสร้างผลลัพธ์ที่มีความแม่นยำและสอดคล้องกับเป้าหมายของงานที่กำหนดไว้

3) การนำเสนอผลลัพธ์ผ่านเว็บไซต์

ผลลัพธ์จากการพยากรณ์ด้วยตัวแบบจะถูกนำมาแสดงในรูปแบบ กราฟเส้น (Line Chart บนเว็บไซต์ โดยใช้ PHP HTML และ Java Script

3.2 การออกแบบเว็บไซต์

3.2.1 การวิเคราะห์และออกแบบฐานข้อมูล

การวิเคราะห์และออกแบบฐานข้อมูลสำหรับระบบวิเคราะห์ข้อมูลและพยากรณ์การใช้น้ำประปา ได้ออกแบบโดยคำนึงถึงประสิทธิภาพในการดำเนินงานและความสามารถในการรองรับการขยายตัวของระบบในอนาคต ซึ่งฟังก์ชันหลักของระบบประกอบด้วย

- การเพิ่มข้อมูล
- การแก้ไข
- การลบ
- การกรองข้อมูลในการค้นหา
- การอัปโหลดข้อมูลในรูปแบบ CSV.

และระบบยังมีฟังก์ชันการ ล็อกอินหรือการเข้าสู่ระบบ เพื่อกำหนดสิทธิ์การเข้าถึงของผู้ดูแลระบบ ซึ่งช่วยให้สามารถจัดการข้อมูลได้อย่างเป็นระบบและปลอดภัย ฐานข้อมูลของระบบถูกออกแบบให้สามารถจัดเก็บข้อมูลที่สำคัญเกี่ยวกับการใช้น้ำประปา ได้แก่

- ปี
- รหัสจังหวัด
- ชื่อจังหวัด
- รหัสอำเภอ
- ชื่ออำเภอ
- รหัสกลุ่มประเภทผู้ใช้
- ชื่อกลุ่มประเภทผู้ใช้
- ปริมาณน้ำ
- ข้อมูลการพยากรณ์แนวโน้มการใช้น้ำประปาในอนาคต

และระบบมีการนำเสนอข้อมูลในรูปแบบที่หลากหลาย เช่น แผนภูมิสถิติ ตารางสรุปผลข้อมูล และแผนภาพการพยากรณ์ โดยระบบมีการกำหนดระดับสิทธิ์ของผู้ใช้งานระบบ ซึ่งสิทธิ์การใช้งานของระบบแบ่งออกเป็น 2 ระดับ ได้แก่

1. ผู้ใช้งานทั่วไป (User)
 - สามารถดูแนวโน้มการใช้น้ำประปา ในแต่ละปีผ่านแผนภูมิและสถิติที่แสดงข้อมูลการใช้น้ำประปาในแต่ละเขตพื้นที่
 - สามารถเข้าถึงข้อมูลสารสนเทศ ในรูปแบบตารางสรุปข้อมูลสถิติการใช้น้ำประปา

- สามารถใช้ระบบกรองข้อมูล เพื่อค้นหาและแสดงข้อมูลในรูปแบบแผนภาพสรุป
- สามารถดูข้อมูลการพยากรณ์แนวโน้มการใช้น้ำประปาในอนาคต

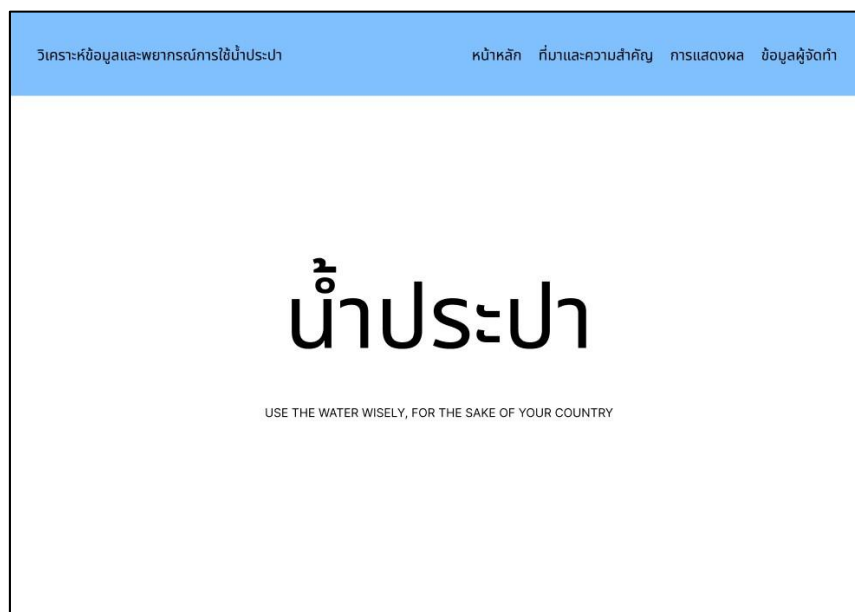
2. ผู้ดูแลระบบ (Admin)

- สามารถเผยแพร่ข้อมูลสารสนเทศ เพื่อให้ผู้ใช้งานทั่วไปเข้าถึงข้อมูลที่ได้รับอนุญาต
- สามารถอัปเดตและอัปเดตข้อมูล เกี่ยวกับการใช้น้ำประปาเข้าสู่ระบบ
- สามารถเพิ่ม ลบ และแก้ไขข้อมูล การใช้น้ำประปาได้ เพื่อให้มั่นใจว่าข้อมูลที่เผยแพร่มีความถูกต้องและเป็นปัจจุบัน

การออกแบบฐานข้อมูลช่วยให้ระบบสามารถทำงานได้อย่างเป็นระบบ มีประสิทธิภาพ และรองรับการจัดเก็บข้อมูลปริมาณมาก และระบบยังคำนึงถึงความปลอดภัยของข้อมูล โดยมีการกำหนดสิทธิ์การเข้าถึงที่ชัดเจนตามระดับของผู้ใช้งาน เพื่อให้มั่นใจว่าข้อมูลที่นำเสนอมีความถูกต้องและน่าเชื่อถือ

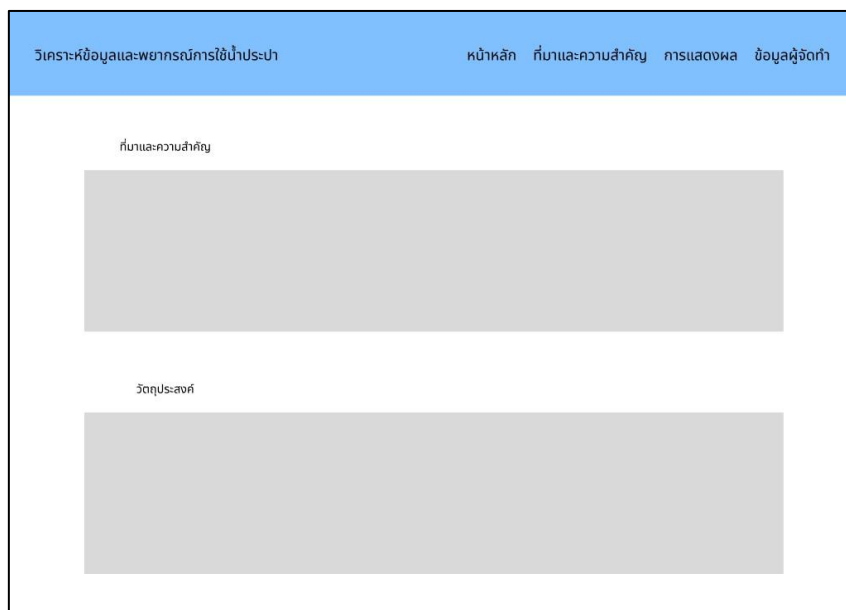
3.2.2 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บ



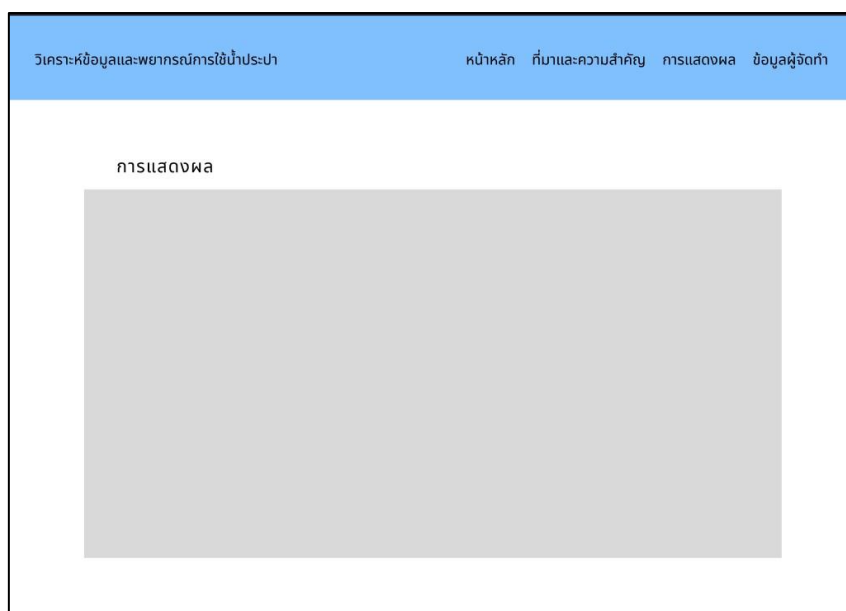
ภาพที่ 3.3 Wireframe หน้าหลัก

2) หน้าที่มาและความสำคัญ วัตถุประสงค์ของโครงการ



ภาพที่ 3.4 Wireframe ที่มาและความสำคัญ

3) หน้าแสดงผล Dashboard ปริมาณการใช้น้ำประปา และการพยากรณ์ ปริมาณการใช้น้ำประปาในอนาคต



ภาพที่ 3.5 Wireframe การแสดงผล

4) หน้าแสดงข้อมูลผู้จัดทำโครงการ



ภาพที่ 3.6 Wireframe ข้อมูลผู้จัดทำ