

# โครงการ

เรื่อง

## The Data Visualization of Suicide Rates Overviews 1985 to 2016

เสนอ

อาจารย์ดร.วนิดา พฤทธิวิทยา

จัดทำโดย

- 1.นายหิรัญย์ ศรีรุ่งเรืองชัย รหัส  
6209610176
- 2.นางสาวสุภัทสร พิมลพันธ์ รหัส  
6209680013
- 3.นางสาวณิชากร สอนเนย รหัส  
6209680849

ภาคเรียนที่ 1 ปีการศึกษา 2563

รายวิชา หลักวิทยาการข้อมูล (Principles of Data  
Science) รหัสวิชา คพ245

**ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และ  
เทคโนโลยี  
มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต**

**บทที่ 1**

**บทนำ**

**(Introduction)**

**1.1 ที่มาและความสำคัญ**

ปัจจุบันสังคมภายในโลกมีอัตราการแข่งขันมากขึ้น ไม่ว่าจะเป็นสังคมระดับวัยเรียนที่มีการแข่งขันกันในเรื่องของผลการเรียน การเป็นที่ยอมรับในสังคม หรือสังคมระดับวัยทำงานที่มีการแข่งขันกันในเรื่องของการทำงาน การเป็นที่ยอมรับของทุกคนในสถานที่ทำงานตั้งแต่เพื่อนร่วมงานไปจนถึงหัวหน้าหรือผู้นำของสถานที่ทำงาน จากการแข่งขันดังที่กล่าวไว้ข้างต้นเป็นสาเหตุหนึ่งของความเครียดที่สามารถพัฒนาต่อเป็นโรคของจิตเวชได้ ซึ่งโรคจิตเวชเป็นสาเหตุหลักที่ทำให้บุคคลเลือกที่จะทำร้ายตัวเองไปจนถึงฆ่าตัวตาย

การฆ่าตัวตายเป็นเหตุการณ์ที่สามารถพบได้จากบุคคลทุกช่วงอายุและทุกประเทศ จากข้อมูลทางสถิติแสดงให้เห็นว่าอัตราการฆ่าตัวตายมีแนวโน้มที่จะสูงมากขึ้นในทุกๆปี ซึ่งหากเป็นประเทศที่มีอัตราการแข่งขันสูงไม่ว่าจะเป็นการแข่งขันภายในประเทศหรือนอกประเทศ อัตราการฆ่าตัวตายจะสูงขึ้นเช่นเดียวกับอัตราการแข่งขัน ดังนั้นทางคณะผู้จัดทำจึงเลือกจัดทำโครงการ "The Data Visualization of Suicide Rates Overviews 1985 - 2016" ขึ้นมา เพื่อที่จะวิเคราะห์ข้อมูลทางสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีของช่วงปีคริสต์ศักราช 1985 ถึง 2016 และแสดงข้อมูลเกี่ยวกับอัตราการฆ่าตัวตายทั่วโลกในช่วงเวลาดัง

กล่าวเพื่อที่จะป้องกันไม่ให้อัตราการฆ่าตัวตายมีแนวโน้มที่จะเพิ่มขึ้นในปีถัดไป

## 1.2 วัตถุประสงค์

1.2.1 เพื่อวิเคราะห์ข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีได้อย่างถูกต้องและครบถ้วน

1.2.2 เพื่อแสดงข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีผ่านการวิเคราะห์อย่างครบถ้วนและถูกต้อง

1.2.3 เพื่อเผยแพร่ข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีเพื่อป้องกันไม่ให้อัตราการฆ่าตัวตายทั่วโลกมีแนวโน้มที่จะสูงขึ้นในปีถัดไป

## 1.3 ขอบเขตของการทำโครงการ

การทำโครงการครั้งนี้ใช้google colab ในการวิเคราะห์และแสดงข้อมูล และใช้ข้อมูลสถิติของการฆ่าตัวตายในแต่ละปีมาจากเว็บไซต์ดังต่อไปนี้:<https://www.kaggle.com/lmorgan95/r-suicide-rates-in-depth-stats-insights> และเว็บไซต์ <https://www.kaggle.com/tavoosi/suicide-data-full-interactive-dashboard>. และทำการเขียนรายงานแสดงข้อมูลบนแอป Google Sheet

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ทราบถึงข้อมูลสถิติของอัตราการฆ่าตัวตายในแต่ละปี

1.4.2 ลดอัตราการฆ่าตัวตายในแต่ละปี

1.4.3 สามารถวิเคราะห์ข้อมูลเกี่ยวกับอัตราการฆ่าตัวตายในแต่ละปี

1.4.4 สามารถแสดงข้อมูลเกี่ยวกับอัตราการฆ่าตัวตายในแต่ละปีได้อย่างชัดเจนและถูกต้อง

## บทที่ 2

### ชุดข้อมูลที่เกี่ยวข้อง (Data Set)

#### 2.1 อภิปรายชุดข้อมูล

ชุดข้อมูลดังกล่าวเป็นชุดข้อมูลของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปี โดยขอบเขตของช่วงเวลาในข้อมูลมีตั้งแต่ปีคริสต์ศักราช 1985 จนถึงปี 2016 ซึ่งมีเว็บไซต์ที่นำข้อมูลในส่วนนี้มาแสดงผลด้วยเช่นกัน ข้อมูลดังกล่าวที่ทางคณะผู้จัดทำค้นคว้ายังมีข้อมูลที่เสียหาย ไม่ครบถ้วนมากพอที่จะสรุปผลเพื่อทำการแสดงข้อมูล ดังนั้นทางคณะผู้จัดทำจึงนำข้อมูลส่วนนี้มาวิเคราะห์ข้อมูล

เพื่อแสดงข้อมูลที่ผ่านการวิเคราะห์จนกระทั่งข้อมูลมีความชัดเจน และถูกต้อง

## 2.2 แหล่งที่มาของชุดข้อมูล

2.2.1 ชุดข้อมูลที่ใช้ในการทำโครงการสืบค้นจากเว็บไซต์ <https://www.kaggle.com/Imorgan95/r-suicide-rates-in-depth-stats-insights> ซึ่งเป็นเว็บไซต์ที่รวบรวมข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปี และเว็บไซต์ <https://www.kaggle.com/tavoosi/suicide-data-full-interactive-dashboard>. ที่แสดงข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีในรูปแบบของแผนภูมิ

## 2.3 ส่วนประกอบของชุดข้อมูล

ส่วนประกอบของข้อมูลมีดังนี้

2.3.1 ข้อมูลของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปี โดยเป็นข้อมูลระยะเวลาภายในปีคริสต์ศักราช 1985 ถึง 2016

2.3.2 ข้อมูลบนเว็บไซต์ที่เกิดขึ้นที่แสดงขึ้นมา ซึ่งสามารถแปลงข้อมูลลงในไฟล์เพื่อที่จะทำข้อมูลมาวิเคราะห์ได้

2.3.3 ข้อมูลบนเว็บไซต์ที่แสดงข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีในลักษณะแผนภูมิ

# บทที่ 3

## การวิเคราะห์ข้อมูล (Methodology)

### 3.1 การเตรียมการประมวลผลและวิเคราะห์ข้อมูล

#### 3.1.1 เครื่องมือที่ใช้

- 1) เว็บไซต์ Google Colab ใช้สำหรับเขียน code ในการวิเคราะห์ข้อมูล แสดงข้อมูล และจัดทำโมเดลข้อมูล
- 2) Library pandas สำหรับนำเข้าข้อมูลจากไฟล์เพื่อวิเคราะห์ข้อมูล และเก็บข้อมูลในรูปแบบ Data frame หรือสร้างแผนภูมิสำหรับ Data Visualization ได้ ตัวอย่างเช่น จาก pandas.plotting ได้ import scatter\_matrix เพื่อที่จะทำการสร้างแผนภูมิการกระจาย
- 3) Library numpy สำหรับคำนวณและวิเคราะห์ข้อมูล
- 4) Library Seaborn สำหรับการทำ Data Visualization
- 5) Library matplotlib.pyplot สำหรับการทำ Data Visualization
- 6) Library sklearn สำหรับใช้ในการสร้างโมเดลจากชุดข้อมูล ตัวอย่างเช่น จาก library sklearn.tree ได้ import library DecisionTreeClassifier เพื่อสร้างโมเดล DecisionTree ขึ้นมา

### **3.2 ขั้นตอนการประมวลผลและวิเคราะห์ข้อมูล**

- 3.2.1 การแสดงสถิติและคุณลักษณะเบื้องต้นของชุดข้อมูล
- 3.2.2 การทำ Data Cleaning
- 3.2.3 การทำ Data Wrangling
- 3.2.4 การใช้เครื่องมือสำหรับ Data Visualization
- 3.2.5 การเลือกใช้โมเดลที่สอดคล้องกับชุดข้อมูล

### **3.3 ผลการดำเนินงาน**

จากการใช้โมเดลที่สอดคล้องกับชุดข้อมูลสามารถอภิปรายผลได้ด้วย Google Colab ซึ่งสามารถดูข้อมูลได้ในลิงค์ Google Colab ดังต่อไปนี้ :

[https://colab.research.google.com/drive/1lr26lCIH3Cl7QF60uy\\_TOHe7p3vqDtmR?authuser=1#scrollTo=DsgZ\\_fXJ2s84](https://colab.research.google.com/drive/1lr26lCIH3Cl7QF60uy_TOHe7p3vqDtmR?authuser=1#scrollTo=DsgZ_fXJ2s84)

## บทที่ 4

### สรุปผลการดำเนินงาน (Conclusion)

#### 4.1 สรุปผล

โครงงานงานนี้เป็นโครงงานวิเคราะห์และแสดงข้อมูลเกี่ยวกับอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีภายในช่วงระยะเวลาคริสต์ศักราช 1985 ถึง 2015 ซึ่งมีแหล่งที่มาจากเว็บไซต์

<https://www>

[.kaggle.com/lmorgan95/r-suicide-rates-in-depth-stats-insights](https://www.kaggle.com/lmorgan95/r-suicide-rates-in-depth-stats-insights) และเว็บไซต์ <https://www>

[.kaggle.com/tavoosi/suicide-data-full-interactive-dashboard](https://www.kaggle.com/tavoosi/suicide-data-full-interactive-dashboard) ใช้ Google colab ในการเขียน Code สำหรับวิเคราะห์และแสดงข้อมูล และทำการเขียนรายงานอภิปรายผลใน Google Sheet

ขั้นตอนการดำเนินงานเริ่มจากการหาแหล่งข้อมูลสำหรับการวิเคราะห์ เมื่อได้แหล่งข้อมูลที่สามารถใช้ได้แล้วจึงนำไฟล์ข้อมูลมาอัปโหลดบน Google Drive เพื่อนำข้อมูลมาแสดงผลบนเว็บไซต์ Google Colab สำหรับนำข้อมูลมาวิเคราะห์และแสดงผล หลังจากนั้นจึงเริ่มทำการเขียน Code เพื่อจัดการข้อมูลที่เสียหรือข้อมูลในส่วนที่ไม่ต้องการ เมื่อได้ข้อมูลครบตามที่ต้องการแล้วจึงเริ่มทำการเรียงอันดับข้อมูลและแสดงข้อมูลออกมาในรูปแบบของแผนภูมิกราฟ เมื่อได้กราฟข้อมูลแล้วจึงเริ่มสร้างโมเดลชุดข้อมูลเพื่ออภิปรายสรุปผลข้อมูล และเริ่มจัดทำตัวรายงานสำหรับสรุปผลและแสดงข้อมูล

ผลการดำเนินเป็นไปตามวัตถุประสงค์กล่าวคือ สามารถวิเคราะห์ข้อมูลและแสดงข้อมูลได้อย่างชัดเจนและถูกต้อง และ

สามารถเผยแพร่ข้อมูลสถิติของอัตราการฆ่าตัวตายทั่วโลกในแต่ละปีภายในช่วงระยะเวลาคริสต์ศักราช 1985 ถึง 2015 เพื่อสร้างความตระหนักเกี่ยวกับอัตราการฆ่าตัวตายที่สูงขึ้นเพื่อป้องกันไม่ให้แนวโน้มที่อัตราการฆ่าตัวตายสูงขึ้นในปีถัดไป

## **4.2 ปัญหาและการแก้ปัญหา**

**4.2.1 ปัญหาด้านการจัดการข้อมูล** เนื่องจากข้อมูลที่ต้องการมีหลายส่วนและต้องอาศัยการทดลองเขียน Code จากในห้องเรียนและจากการค้นคว้าจึงทำให้ข้อมูลที่แสดงไม่ตรงกับที่ต้องการหรือไม่ถูกต้อง มีวิธีแก้ไขคือ การแก้ไข Code ในส่วนของการจัดการข้อมูลให้ได้ข้อมูลที่ต้องการ

**4.2.2 ปัญหาด้านเวลาในการทำโครงงาน** เนื่องจากสมาชิกภายในกลุ่มมีเวลาว่างไม่ตรงกัน ทำให้นัดหมายทำโครงงานได้ยาก มีวิธีแก้ไขคือ ทำการนัดเวลาที่แน่นอนสำหรับการทำงาน หรือแบ่งงานภายในกลุ่มแล้วแยกย้ายกันทำโครงงาน

**4.2.3 ปัญหาด้านปริมาณของข้อมูล** เนื่องจากปริมาณของข้อมูลที่มีมาก ทำให้การจัดการข้อมูลให้มีความถูกต้อง (Accuracy) น้อย มีวิธีแก้ไขคือ กำหนดช่วงของข้อมูลให้ชัดเจน และรวมกลุ่มข้อมูล

## **4.3 การนำไปใช้ประโยชน์/แนวทางการประยุกต์หรือพัฒนาต่อยอดในลักษณะอื่นๆ**

### **4.3.1 การนำไปใช้ประโยชน์**

- 1) สามารถวิเคราะห์ข้อมูลได้อย่างชัดเจนและถูกต้อง
- 2) สามารถแสดงข้อมูลได้อย่างชัดเจนและถูกต้อง
- 3) สามารถนำชุดข้อมูลที่ผ่านการวิเคราะห์แล้วไปเผยแพร่เพื่อเป็นข้อมูลแก่บุคคลที่สนใจเกี่ยวกับข้อมูลดังกล่าวได้

### **4.3.2 แนวทางการประยุกต์หรือพัฒนาต่อยอดในลักษณะอื่นๆ**



- 1) นำความรู้ที่ได้จากการวิเคราะห์และแสดงข้อมูลไปใช้ในการวิเคราะห์และแสดงข้อมูลในครั้งถัดไปในชุดข้อมูลอื่นได้
- 2) หากสนใจที่จะศึกษาเกี่ยวกับData Science ผู้อ่านสามารถนำความรู้ที่ได้จากการวิเคราะห์และแสดงข้อมูลไปศึกษาเพื่อต่อยอดองค์ความรู้ได้
- 3) สามารถประยุกต์ปัญหาจากการทำโครงการนี้ไปใช้กับการทำงานอื่นได้