**การขายค้าประเวณีของไทยบนทวิตเตอร์**

**Thai Prostitution on Twitter**

**สุรเดช อินทกรณ์1 และ ภัททิรา ไทยทอแสง2**

1คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน

2คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล

Emails: suradej.i@ku.th, pattira.tha@mahidol.ac.th

**บทคัดย่อ**

ทวิตเตอร์กำลังเป็นสื่อสังคมออนไลน์ที่มีความนิยมในประเทศไทย ผู้ใช้สามารถสร้างข้อความขนาดสั้นที่เรียกว่า “ทวีต” ทวีตสามารถมีข้อมูลอื่นนอกเหนือจากข้อความ เช่น พิกัดภูมิศาสตร์ของผู้ใช้ ซึ่งบางบริษัทใช้ข้อมูลนี้ในการโฆษณา การขายบริการทางเพศเป็นธุรกิจที่มีการโฆษณาผ่านทวิตเตอร์เช่นกัน ด้วยสถานะผิดกฎหมายของการขายบริการทางเพศในประเทศไทย จึงยังไม่มีสถิติผู้ขายบริการทางเพศแบบเป็นทางการ งานวิจัยชิ้นนี้มีจุดประสงค์เพื่อศึกษาความสัมพันธ์เชิงสถิติของตัวแปรทางสังคมต่อจำนวนผู้ค้าบริการทางเพศในแต่ละจังหวัด ข้อมูลทวิตเตอร์ก่อนการประมวลผลที่เก็บระหว่าง วันที่ 30 มิถุนายน ถึง 3 พฤศจิกายน 2562 เป็นจำนวน 4.8 ล้านทวีต ข้อมูลที่เกี่ยวข้องกับการค้าบริการทางเพศได้จำแนกออกมาโดยใช้คำสำคัญ พิกัดภูมิศาสตร์ของทวีตเปลี่ยนเป็นชื่อจังหวัดโดยกระบวนการ “แปลงพิกัดเป็นที่อยู่จริง” จังหวัดที่มีจำนวนผู้ทวีตเกี่ยวกับการขายบริการทางเพศ สูงสุด 5 อันดับแรกคือ กรุงเทพมหานคร นครราชสีมา ระยอง ชลบุรี และนนทบุรี การวิเคราะห์สหสัมพันธ์เพียร์สันระหว่างจำนวนผู้ทวีตเกี่ยวกับการขายบริการทางเพศกับจำนวนประชากรในแต่ละจังหวัด ผลิตภัณฑ์มวลรวมจังหวัดและผลิตภัณฑ์มวลรวมจังหวัดต่อคน พบว่ามีค่าเป็น 0.9224, 0.9379, และ 0.4526 ตามลำดับ สรุปได้ว่าขนาดผลิตภัณฑ์มวลรวมจังหวัดมีความสัมพันธ์เชิงสถิติต่อการขายบริการทางเพศในสื่อสังคมออนไลน์มากกว่าผลิตภัณฑ์มวลรวมจังหวัดต่อคน ดังนั้นขนาดเศรษฐกิจของจังหวัดจึงมีความเกี่ยวข้องอย่างมากกับจำนวนผู้ค้าบริการทางเพศบนทวิตเตอร์

**Abstract**

Recently, Twitter has become one of the most popular social media platforms among Thai netizens. Users can create a short message called “tweet.” Tweets can also contain other forms of information such as locations, which some companies use them to target their advertisements. Prostitution is advertised on twitter as well. Due to the illegal status of prostitution in Thailand, there is no official number of prostitutes. However, our research provides a way towards the estimation. This research aims to study the characteristics of online prostitution, such as location, frequency, type of service, and content. We have collected tweets with indicated locations to represent the number of prostitutes in each province of Thailand. Python and Tweepy were used in the process of collecting data. The number of tweets collected between June 30, 2019 and November 3, 2019 is approximately 4.8 million. Prostitution-related tweets were extracted by means of keywords. Tweet locations were converted to Thai provinces by “reverse geocoding” method to count the number of prostitutes. The result indicates that provinces with the most prostitution-related tweets are Bangkok, Nakhon Ratchasima, Chon Buri, Rayong, and Nonthaburi, consecutively. The Pearson correlation coefficients between the number of prostitutes and other three variables (population, gross domestic product, and gross domestic product per capita) are 0.9224, 0.9379, and 0.4526, respectively. In conclusion, the statistical relationship between the gross domestic product (GDP) and the number of prostitutes is stronger than that of the GDP per capita and the number of prostitutes. Thus, the scale of economy is effectively related to the number of prostitutes on Twitter.

**1． บทนำ**

สื่อสังคมออนไลน์ใช้ในการวิเคราะห์ด้านสังคมศาสตร์อย่างกว้างขวาง ทวิตเตอร์เป็นหนึ่งในสื่อสังคมออนไลน์ที่นิยมใช้ในการวิจัยทั้งด้านธุรกิจและภาครัฐเพื่อวิเคราะห์พฤติกรรมของมนุษย์ (Pfeffer, Mayer and Morstatter, 2018) ในงานวิจัยชิ้นนี้ คณะผู้วิจัยมีจุดประสงค์ในการวิเคราะห์ข้อมูลทางสถิติของการขายบริการทางเพศบนสื่อสังคมออนไลน์

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยชิ้นนี้คือ

1. นำเสนอความสถิติการค้าบริการทางเพศบนสื่อสังคมออนไลน์ในแต่ละจังหวัดในประเทศไทย
2. แสดงให้เห็นว่าจำนวนการค้าบริการทางเพศบนสื่อสังคมออนไลน์มีความสัมพันธ์เชิงสถิติต่อขนาดเศรษฐกิจโดยรวมมากกว่าต่อจำนวนประชากร
3. อภิปรายเชิงคุณภาพเกี่ยวกับคุณลักษณะของข้อมูลการค้าบริการทางเพศบนสื่อสังคมออนไลน์

**2. งานวิจัยที่เกี่ยวข้อง**

ในประเทศไทย ธุรกรรมการค้าประเวณีจัดเป็นเศรษฐกิจใต้ดิน เนื่องจากเป็นสิ่งผิดกฎหมาย มูลเหตุที่บุคคลตัดสินใจเข้าสู่อาชีพค้าประเวณีนั้น ผลการศึกษาปัจจุบันส่วนใหญ่ชี้ไปในทางเดียวกันคือ เหตุผลด้านเศรษฐกิจ เช่น หาเงินง่าย มูลค่าทางเศรษฐกิจในการลอบค้าประเวณีในสถานบริการนวดแผนโบราณจาก 13 จังหวัดสูงถึง 835,615,000 บาท ถึง 1,671,230,000บาท ต่อปี โดย 13 จังหวัดประกอบด้วย จังหวัดที่มีผลิตภัณฑ์มวลรวมต่อหัวสูงที่สุด 10 อันดับแรกของประเทศ ตามด้วยเชียงใหม่ นนทบุรี และนครราชสีมา (ภูวาเดช โหราเรือง, 2018) การใช้ทวิตเตอร์ในงานวิจัยปรากฏการณ์ทางสังคมไทย ส่วนใหญ่จะเป็นงานเชิงคุณภาพ เช่น การใช้ทวิตเตอร์ของผู้สื่อข่าวโทรทัศน์ (กนิษฐา ไชยแสง และ ขจร ฝ้ายเทศ, 2014) หรือ การนำเสนอภาพลักษณ์ของ “เกย์” และ “เลสเบี้ยน” ใน ทวิตเตอร์ (พิมพ์วิภา มีงามดี, สาธนี นาแซง, และ พิชญาวี ทองกลาง, 2020) ในขณะที่บางส่วนเป็นงานเชิงปริมาณเช่น การวิเคราะห์ทวิตเตอร์ในเหตุการณ์สวรรคตของพระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดช (อารยา พุดตาล, 2017 ) อย่างไรก็ตาม การวิเคราะห์เชิงปริมาณส่วนใหญ่เป็นการวิเคราะห์แบบง่าย ๆ เช่นความถี่ของการเกิดคำสำคัญ ส่วนการวิเคราะห์แบบอัตโนมัติขั้นสูงกว่านั้น เช่น ความรู้สึกของผู้ทวีต จำเป็นต้องใช้วิธีการประมวลผลด้านภาษาไทยขั้นสูง ตัวอย่างเช่น ในงานของ (Vateekul and Koomsubha, 2016) โดยพบว่าการใช้วิธีการเรียนรู้ทางเครื่องจักรที่เรียกว่า Long Short Term Memory (LSTM) ได้ผลดีที่สุด

ตำแหน่งทางภูมิศาสตร์เป็นส่วนสำคัญในทวีตเช่นกัน [Graham](https://www.tandfonline.com/author/Graham%2C+Mark) ศึกษาความสัมพันธ์ระหว่างภาษาและตำแหน่งทางภูมิศาสตร์บนทวิตเตอร์ การอนุมานตำแหน่งจากเพียงข้อความนั้นยากมาก ดังนั้นนักวิจัยจำนวนมากจึงเลือกใช้ตำแหน่งของทวีตที่สร้างจากอุปกรณ์บอกตำแหน่งเช่น GPS ในสมาร์ทโฟน อย่างไรก็ตามขนาดของชุดข้อมูลจะน้อยลงมาก เนื่องจากมีเพียงแค่ 0.7% ที่มีข้อมูลตำแหน่งทางภูมิศาสตร์จากอุปกรณ์บอกตำแหน่ง (Graham, Hale and Gaffney, 2014) ข้อมูลทวิตเตอร์โดยมากจะเก็บผ่าน Twitter Application Programming Interface (Twitter API) โดยทวิตเตอร์จะสุ่มกลุ่มตัวอย่างปริมาณ 1 % แบบไม่เสียค่าใช้จ่ายและ 10% หากเสียค่าใช้จ่าย ทวิตเตอร์ไม่เปิดเผยวิธีการสุ่มต่อสาธารณะ โดยทั่วไปอาจอนุมานได้ว่าข้อมูลนี้เป็นการสุ่มอย่างง่าย (random sampling) อย่างไรก็ตาม Pfeffer ได้ใช้วิธีการวิศวกรรมย้อนกลับและพบว่านักวิจัยไม่ควรตั้งสมมติฐานว่า ข้อมูลที่ได้จากทวิตเตอร์เป็นการสุ่มกลุ่มตัวอย่างแบบ random sampling (Pfeffer, Mayer and Morstatter, 2018)

1. **วิธีการทดลอง**

**3.1 การเก็บข้อมูล**

คณะผู้วิจัยเก็บข้อมูลผ่าน Twitter Streaming API แบบไม่เสียเงิน โดยกำหนดให้เก็บข้อมูลเฉพาะทวีตที่มีข้อมูลตำแหน่งภูมิศาสตร์ กล่องขอบเขต (Bounding Box) ที่ใช้เป็นพารามิเตอร์ในการค้นหาจะครอบคลุมเฉพาะประเทศไทยเท่านั้น คณะผู้วิจัยสร้างโค้ดที่ใช้ในการเก็บข้อมูลด้วยภาษาไพธอน โดยใช้ไลบรารีทวีพาย[[1]](#footnote-1)เพื่อช่วยให้การพัฒนาโค้ดง่ายขึ้น โค้ดที่เขียนขึ้นเก็บข้อมูลผ่านการสตรีมมิ่งในช่วงวันที่ 30 มิถุนายน ถึง 3 พฤศจิกายน 2562   
ข้อมูลที่สตรีมมิ่งเข้ามาจะเก็บในรูปทวีตออบเจค และเก็บไว้ในลิส (List) รอจนกระทั่งเก็บข้อมูลได้ครบ 1000 ทวีต และจะบันทึกลงฮาร์ดดิส ทั้งนี้เพื่อเพิ่มประสิทธิภาพของอินพุตและเอาท์พุต (IO) รูปแบบของไฟล์จะเป็นการทำออบเจคซีเรียลไลเซชัน (Object Serialization) ลิสของทวีตออบเจค โดยพิคเคิล[[2]](#footnote-2)(Pickle) ไลบรารีใช้ในการทำออบเจคซีเรียลไลเซชัน

ทวีตที่ใช้ในการวิจัยนี้ไม่ใช่ทวีตทั้งหมดอย่างแท้จริงเนื่องจากทวิตเตอร์สตรีมมิ่งจะสุ่มทวีตมา 1% สำหรับแบบไม่เสียเงิน และ 10% สำหรับแบบเสียเงิน คณะผู้วิจัยใช้ในรูปแบบไม่เสียเงิน ดังนั้นกลุ่มตัวอย่างจะเป็น 1% ของจำนวนทวีตจริง อย่างไรก็ตาม กลุ่มตัวอย่างนี้อาจจะไม่ได้เป็นการสุ่ม แบบ random sampling อย่างแท้จริง ดังที่ Pfeffer ได้ทดลองไว้ นอกจากนี้ข้อมูลเก็บเฉพาะทวีตที่มีพิกัดภูมิศาสตร์อยู่ในประเทศไทยเท่านั้น นั่นคือผู้ใช้ต้องใช้อุปกรณ์ที่สามารถระบุตำแหน่งภูมิศาสตร์ได้ (เช่นสมาร์ทโฟนที่มี GPS) และอนุญาตให้ระบบเพิ่มข้อมูลตำแหน่งภูมิศาสตร์ในทวีตด้วย หากผู้ใช้ทวีตข้อความเกี่ยวกับการค้าประเวณีแต่ไม่ได้อนุญาตให้ระบบเพิ่มข้อมูลตำแหน่งภูมิศาสตร์ในทวีต ข้อมูลทวีตนั้นจะไม่ได้อยู่ในชุดข้อมูลนี้ ดังนั้น ข้อมูลชุดนี้อาจมีสาเหตุที่ทำให้กลุ่มตัวอย่างไม่ได้ เป็นแบบ random sampling ของการใช้ทวิตเตอร์โดยผู้ค้าประเวณีอย่างแท้จริงอยู่สองประการคือ วิธีการสุ่มของทวิตเตอร์ และ การเลือกเฉพาะข้อมูลทวิตเตอร์ที่มีพิกัดภูมิศาสตร์ อย่างไรก็ตามนี่เป็นวิธีการที่ดีที่สุดในความเห็นของคณะผู้วิจัย สำหรับวิธีการสุ่มของทวิตเตอร์ คณะผู้วิจัยไม่สามารถแก้ไขได้เนื่องจากเป็นข้อกำหนดจากทวิตเตอร์ และคณะผู้วิจัยต้องการใช้ข้อมูลทวีตที่มีพิกัดภูมิศาสตร์เท่านั้นจึงเลือกเฉพาะข้อมูลที่มีพิกัดภูมิศาสตร์อยู่ในไทย ในช่วงเวลาประมาณ 3 เดือน คณะผู้วิจัยเก็บข้อมูลได้ทั้งหมด 4,875,780 ทวีต สาเหตุที่ปริมาณทวีตมีเพียงแค่ประมาณ 4 ล้านเนื่องจากวิธีการสุ่มเพียง 1% นอกจากนั้นเรายังกรองข้อมูลที่ไม่มีข้อมูลพิกัดภูมิศาสตร์ด้วย และทวีตเพียง 0.7% เท่านั้นที่มีข้อมูลพิกัดภูมิศาสตร์ (Graham,Hale and Gaffney,2014)

**3.2 รายละเอียดข้อมูล**

ทวีตออบเจคประกอบด้วยข้อมูลหลายส่วน นอกเหนือจากข้อความขนาด 280 ตัวอักษรแล้ว ยังมีข้อมูลอื่น ๆ สามารถใช้ในการวิเคราะห์ได้ เช่น วันเวลา อุปกรณ์ที่ใช้ (เช่น iphone หรือ android) พิกัดภูมิศาสตร์ ข้อมูลของผู้โพส เป็นการตอบกลับไปที่ทวีตใด ฯลฯ ตารางที่ 1 ด้านล่างเป็นตัวอย่างข้อมูลบางส่วนใน tweet โดยรายละเอียดที่สมบูรณ์สามารถอ่านได้จาก เอกสารของ tweepy และ twitter streaming api[[3]](#footnote-3) ข้อมูลตัวอย่างในตาราง คณะผู้วิจัยตัดมาเพียงบางส่วนเพื่อปกป้องความเป็นส่วนตัวของผู้ใช้ คำหยาบคายบางส่วนก็ละไปจากตัวอย่างข้อมูลด้วย

ตารางที่ 1

|  |  |  |
| --- | --- | --- |
| ชื่อคุณลักษณะ | ความหมาย | ตัวอย่าง |
| text | ข้อความที่โพส | งานหาค่าเทอมจ้า  สนใจทักจ้า #ไซด์ไลน์โคราช line: |
| created\_at | วันเวลา | Sun Jun 30 09:29:48 +0000 2019 |
| source | อุปกรณ์ที่ใช้ในการ  โพสทวีต | '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>' |
| in\_reply\_to\_status\_id | ตอบกลับไปยังทวีตใด | None |
| in\_reply\_to\_user\_id | ตอบกลับไปยังผู้ใช้คนใด | None |
| User.description | คำที่ผู้ใช้อธิบายตนเอง | #รับงาน #ไซด์ไลน์ #BDSM #เรียลหญิง #ยืมเมจfosrex |
| User.location | ที่อยู่ของผู้ใช้ | None |
| Place.place\_type | ประเภทของสถานที่  ที่ทวีต | admin |
| Place.name | ชื่อสถานที่ของตำแหน่ง  ที่ทวีต | อ.เมืองนครราชสีมา |
| Place. full\_name | ชื่อสถานที่แบบเต็ม  ที่ทวีต | อ.เมืองนครราชสีมา, จ.นครราชสีมา |
| Place.BoundingBox  .coordinate | พิกัดภูมิศาสตร์ของฺ BoudingBox ที่ทวีต | [[[101.932499, 14.789611], [101.932499, 15.14049], [102.230817, 15.14049], [102.230817, 14.789611]]] |
| extended\_tweet  .entities.hashtags | แฮชแท็กในทวีตพร้อมตำแหน่งตัวอักษรในทวีต | '{'text': 'รับงานโคราช', 'indices': [57, 69]},  {'text': 'ไซด์ไลน์โคราช', 'indices': [84, 98]} |

ส่วนข้อมูลสถานที่ ทวิตเตอร์ให้ข้อมูลที่หลากหลายเช่น ประเภทสถานที่ (เมือง,เขตการปกครอง, เขต,ตำบล หรือ landmark) , ชื่อสถานที่ (Place.name)และ ตำแหน่งของภูมิศาสตร์ของสถานที่ อย่างไรก็ตาม จะพบว่าบางทวีตตำแหน่งนี้ไม่ได้ประกอบด้วยคู่อันดับ ละติจูด กับ ลองจิจูด เพียงหนึ่งคู่ เช่นข้อมูล [[[101.932499, 14.789611], [101.932499, 15.14049], [102.230817, 15.14049], [102.230817, 14.789611] จะพบว่ามีคู่อันดับ ละติจูด กับ ลองจิจูด สี่คู่ โดยในตัวอย่างนี้จะเป็นตัวอย่างของข้อมูลที่เรียกว่าbounding box นั่นคือการกำหนดขอบเขตสถานที่แบบหยาบๆโดยกล่องสี่เหลี่ยม (ในความเป็นจริงขอบเขตของสถานที่ต้องเป็นโพลีกอน)ในตัวอย่างนี้ คู่อันดับ (101.932499, 14.789611)(จะเป็นตัวแทนของมุมด้านซ้ายล่างของสี่เหลี่ยม คู่อันดับ(101.932499, 15.14049)จะเป็นมุมด้านขวาล่างของสี่เหลี่ยม คู่อันดับ (102.230817, 15.14049)จะเป็นมุมด้านขวาบนของสี่เหลี่ยม

ข้อมูลที่ใช้ในการวิจัยนี้ประกอบด้วย ข้อความ และ สถานที่ที่ใช้โพสเป็นหลักเท่านั้น จากตารางที่ 1 จะเห็นว่าผู้ใช้จะสามารถกำหนดตำแหน่งที่อยู่ของตัวเองได้ด้วย(User.location) และผู้ใช้สามารถเลือกที่จะไม่กำหนดได้โดยหากผู้ใช้ไม่กำหนดจะมีค่าเป็น None ข้อมูลบางส่วนแม้ยังไม่ได้ใช้ในการวิจัยนี้ แต่คณะผู้วิจัยพบว่าอาจจะมีประโยชน์ต่อการทำเหมืองข้อมูล (Data Mining) ในอนาคต เช่น คำอธิบายตนเองของผู้ใช้(User.description) พบว่าจากในตัวอย่างผู้ใช้ที่ค้าประเวณีบางคนจะใส่รายละเอียดการค้าบริการไว้ในส่วน (User.description) เลยนอกจากนี้การโต้ตอบกับผู้ใช้อื่นก็อาจนำมาใช้ประโยชน์ในการวิจัยเช่นกัน (in\_reply\_to\_user\_id)

สำหรับข้อมูลสถานที่ ทวิตเตอร์ให้ข้อมูลที่หลากหลาย เช่น ประเภทสถานที่ (เมือง เขตการปกครอง เขต ตำบล หรือ landmark) ชื่อสถานที่ (Place.name) และตำแหน่งของภูมิศาสตร์ของสถานที่ อย่างไรก็ตาม จะพบว่าบางทวีตตำแหน่งนี้ไม่ได้ประกอบด้วยคู่อันดับละติจูดกับลองจิจูดเพียงหนึ่งคู่ เช่น [[[101.932499, 14.789611], [101.932499, 15.14049], [102.230817, 15.14049], [102.230817, 14.789611] จะพบว่ามีคู่อันดับละติจูดกับ ลองจิจูดสี่คู่ โดยในตัวอย่างนี้จะเป็นตัวอย่างของข้อมูลที่เรียกว่าbounding box นั่นคือการกำหนดขอบเขตสถานที่แบบหยาบ ๆ โดยกล่องสี่เหลี่ยม (ในความเป็นจริงขอบเขตของสถานที่ต้องเป็นโพลีกอน) ในตัวอย่างนี้ คู่อันดับ (101.932499, 14.789611) จะเป็นตัวแทนของมุมด้านซ้ายล่างของสี่เหลี่ยม คู่อันดับ (101.932499, 15.14049)จะเป็นมุมด้านขวาล่างของสี่เหลี่ยม คู่อันดับ (102.230817, 15.14049) จะเป็นมุมด้านขวาบนของสี่เหลี่ยม อ้างอิงจาก Graham ข้อมูลพิกัดภูมิศาสตร์สามารถเป็นจุดตำแหน่ง หรือ คู่อันดับละติจูดกับลองจิจูดเพียงหนึ่งคู่ หรือจะเป็นตำแหน่งโดยประมาณที่เป็นกล่องสี่เหลี่ยมก็ได้ ขึ้นอยู่กับการตั้งค่าความเป็นส่วนตัวของผู้ใช้ (User Privacy Setting) (Graham,Hale and Gaffney,2014)

คุณลักษณะที่สำคัญอีกประการของทวีตคือแฮชแท็ก (hashtag) ทวิตเตอร์สตรีมมิ่งได้ประมวลผลแฮชแท็กไว้ให้ผู้ใช้เลย นักพัฒนาสามารถนำแฮชแท็กไปใช้ได้เลยโดยที่ไม่ต้องประมวลผลข้อความซ้ำอีกรอบ นั่นช่วยให้นักพัฒนาไม่จำเป็นต้องเสียเวลาเขียนโค้ดในการแยกแฮชแท็กออกมาซ้ำ ตัวอย่างข้อมูลคือ'{'text': 'รับงานโคราช', 'indices': [57, 69]}, {'text': 'ไซด์ไลน์โคราช', 'indices': [84, 98]} สัญลักษณ์เหล่านี้สื่อความหมายว่ามีแฮชแท็ก #รับงานโคราช อยู่ที่ตัวอักษรตำแหน่งที่ 57 ถึง 69 และ #ไซด์ไลน์โคราช อยู่ที่ตัวอักษรตำแหน่งที่ 84 ถึง 98 การใช้แฮชแท็กจะช่วยให้สืบค้นได้ง่ายขึ้น เช่น ถ้านักพัฒนาต้องการสืบค้นทวีตที่เกี่ยวข้องกับ #ไซด์ไลน์ ก็สามารถกำหนดเงื่อนไขค้นหาให้ทวีตมีแฮชแท็ก #ไซด์ไลน์

**3.3 การจำแนกจังหวัดของทวีต**

คณะผู้วิจัยใช้ข้อมูลพิกัดภูมิศาสตร์ของทวีตในการหาว่าทวีตนั้นโพสจากจังหวัดใด ในกรณีที่ข้อมูลอยู่ในรูปฺ Bounding Box เช่น [101.932499, 14.789611], [101.932499, 15.14049], [102.230817, 15.14049], [102.230817, 14.789611] จุดศูนย์กลางของกล่องจะถูกใช้เป็นตัวแทนของตำแหน่ง ตำแหน่งทางภูมิศาสตร์นี้ (คู่อันดับละติจูดกับลองจิจูด) จะถูกแปลงเป็นชื่อจังหวัดโดยกระบวนการที่เรียกว่า “การแปลงพิกัดเป็นที่อยู่จริง”(reverse geocoding) สาเหตุที่ใช้ตำแหน่งทางภูมิศาสตร์นี้แทนที่จะใช้ข้อมูลชื่อสถานที่แบบเต็มเช่น “อ.เมืองนครราชสีมา, จ.นครราชสีมา” เนื่องจากการประมวลผลทางข้อความมีโอกาสผิดพลาดมากกว่า นอกจากนี้บางทวีตไม่มีข้อมูลจังหวัดเช่น “สามเสนนอก, ประเทศไทย” นอกจากนี้ ชื่อสถานที่ยังมีโอกาสเป็นทั้งภาษาไทยและอังกฤษเช่น “Amphoe Bang Bua Thong, Changwat Nonthaburi” ด้วยเหตุผลดังกล่าวจึงทำให้การประมวลทางข้อความยากกว่ากระบวนการ reverse geocoding

อีกสองแหล่งข้อมูลที่อาจนำมาใช้เพื่อระบุตำแหน่งของทวีตนั้นคือในข้อความของทวีตเอง คณะผู้วิจัยพบว่าทวีตที่เกี่ยวข้องกับการขายบริการทางเพศ มักจะระบุพื้นที่ให้บริการ นอกจากนี้ที่อยู่ของผู้ใช้ (User.location　หรือ profile location) ก็สามารถนำมาใช้ได้ อย่างไรก็ตาม ทั้งสองแหล่งข้อมูลนี้มีความลำบากเชิงเทคนิคในการประมวลผลแบบอัตโนมัติ เนื่องจากต้องสกัดจากข้อความ นอกจากนี้ผู้ใช้อาจไม่ได้ใส่ตำแหน่งที่อยู่จริงอีกด้วย อ้างอิงจาก Graham ข้อมูลในส่วนนี้หาตำแหน่งพิกัดภูมิศาสตร์ได้ยาก เนื่องจากผู้ใช้อาจใส่ข้อมูลเป็นอะไรก็ได้ เช่น Middle-earth ดังนั้นข้อความในส่วนนี้มักมีความผิดพลาดสูง ที่อยู่ของผู้ใช้ (profile location) เป็นข้อความระบุตำแหน่งของตนเองแบบปลายเปิด หากต้องการใช้ข้อมูลในส่วนนี้ จำเป็นต้องใช้แรงงานคนจำนวนมากในการแปลงเป็นพิกัดภูมิศาสตร์อย่างถูกต้อง (เช่นในงาน Takhteyev, Gruzd, and Wellman, 2012) (Graham,Hale and Gaffney, 2014)

**3.4 การจำแนกข้อมูลที่เกี่ยวข้องกับการค้าประเวณี**

การจำแนกข้อมูลที่เกี่ยวข้องกับการค้าประเวณีเป็นขั้นตอนสำคัญมากในงานชิ้นนี้ จุดประสงค์ของขั้นตอนนี้คือแยกทวีตที่เกี่ยวข้องกับการค้าประเวณีออกมาจากทวีตทั้งหมด ความยากลำบากในขั้นตอนนี้คือ ปริมาณข้อมูลจำนวนมากของทวิตเตอร์ การใช้คนมาตัดสินว่าทวีตนี้เกี่ยวข้องกับการค้าประเวณีหรือไม่นั้นต้องใช้ทรัพยากรคนและเวลาสูงมาก ดังนั้นการใช้วิธีการอัตโนมัติจึงเป็นวิธีที่จะช่วยจัดการกับข้อมูลขนาดใหญ่นี้ได้ดีกว่าในทางปฏิบัติ การจำแนกข้อมูลประเภทข้อความโดยอัตโนมัตินั้นค่อนข้างยาก โดยมีปัญหาย่อยที่ต้องแก้จำนวนมากเช่น ปัญหาขอบเขตของคำ นอกจากนั้นบริบทของข้อความก็ค่อนข้างสำคัญ เช่น อาจเป็นการพูดเล่นกับเพื่อน ในงานวิจัยชิ้นนี้ คณะผู้วิจัยใช้วิธีกึ่งอัตโนมัติในการจำแนกข้อมูล ด้วยคำสำคัญสามคำ คือ 'ไซด์ไลน์' 'ขายตัว' และ 'ขายน้ำ' อย่างไรก็ตาม คณะผู้วิจัยพบว่ายังมีข้อผิดพลาดอยู่จึงกรองซ้ำอีกครั้งด้วยมนุษย์ วิธีการนี้แม้จะค่อนข้างแม่นยำแต่ก็จะขาดความสมบูรณ์และไม่สามารถทำกับข้อมูลที่ใหญ่กว่านี้มาก ๆ ได้ วิธีการนี้ไม่สมบูรณ์เนื่องจากขึ้นกับคำสำคัญที่ใช้ แน่นอนว่าทวีตที่เกี่ยวข้องกับการค้าประเวณีอาจใช้คำสำคัญอื่นนอกเหนือจาก 'ไซด์ไลน์' 'ขายตัว' และ 'ขายน้ำ' ดังนั้นนี่จึงเป็นเหตุของความไม่สมบูรณ์ สาเหตุที่ทำกับข้อมูลขนาดที่ใหญ่กว่านี้มาก ๆ ไม่ได้ เนื่องจากการใช้มนุษย์ในการตัดสินจำเป็นต้องใช้ทรัพยากรจำนวนมาก

คณะผู้วิจัยได้ทดลองกรองข้อมูลด้วยคำสำคัญ 2 วิธี คือ กรองจากคำในข้อความ และ กรองด้วยแฮชแท็ก สมมติฐานเบื้องต้นคือการกรองด้วย แฮชแท็ก จะมีความแม่นยำกว่า อย่างไรก็ตามการใช้แฮชแท็กนั้นมีความหลากหลายมากในหัวข้อการค้าประเวณี กล่าวคือ หากผู้ใช้กำหนดแฮชแท็กเกี่ยวกับไซด์ไลน์มักจะตามด้วยชื่อสถานที่เช่น “ทักได้นะค่ะ #รับงาน***ไซด์ไลน์***สุราษฎร์ธานี #รับงานสุราด” จะพบว่าในกรณีนี้ หากกรองด้วยแฮชแท็ก “ไซด์ไลน์” จะไม่สามารถสืบค้นข้อมูลนี้ออกมาได้ในขณะที่หากให้วิธีการค้นหาคำในข้อความจะสามารถสืบค้นได้โดยจะพบที่ตำแหน่งที่เป็นตัวหนาเอียงในตัวอย่างด้านบน อย่างไรก็ตาม การกรองจากคำในข้อความมีโอกาสผิดพลาดมากกว่า เช่น “ไม่มีใครขายชาติได้ ถ้าชาติไม่ขายตัวเอง“ หรือ “ไอ้ตี่ท่ขายน้ำเต้าหู้ในวันนั้นสู่การเป็นเจ้าของสวนกาแฟในวันนี้” จะพบว่าสองทวีตนี้เมื่อค้นด้วยคำสำคัญ ขายตัว และ ขายน้ำ จะพบหลุดการกรองออกมา ในกรณีนี้ หากใช้การตัดคำ (Word Tokenization) ในวิธีการประมวลผลภาษาธรรมชาติ (Natural Language Processing) อาจจะช่วยบรรเทาปัญหาลงได้บ้างเนื่องจากจะสามารถแยกกรณีของข้อความที่ใช้คำว่า ขายน้ำ กับ ขายน้ำเต้าหู้ ได้

กรณีผิดพลาดถัดไปคือการพูดเล่น ในกรณีนี้อาจต้องใช้บริบทอื่น ๆ ประกอบ เช่น ข้อมูลของผู้ใช้ การโต้ตอบกับผู้ใช้คนอื่น ตัวอย่างของกรณีนี้คือ “อยากขายตัวแต่กลัวเขามาขอแบ่งซื้อเป็นกิโล” หรือ “ขายของไม่ได้ก้อต้องขายตัวสะมั้งกู” จะพบว่าผู้ใช้พูดถึงการขายตัวในเชิงพูดเล่น และยังใช้ในบริบททางการเมือง “ไปแก้ผิดจุดอ่ะ กลัวสส.ขายตัว มันต้องแก้ทางการเมือง ครั้งหน้าปชช.ก็ไม่เลือกเอง” สองกรณีที่กล่าวข้างต้นหากใช้การค้นหาโดยแฮชแท็กก็จะพอบรรเทาลงได้เนื่องจากแฮชแท็ก #ขายตัว มักใช้ในกรณีของการค้าประเวณีเป็นหลัก

ในข้อมูลชุดนี้ มีเหตุการณ์พิเศษเกิดเมื่อวันที่ 4 สิงหาคม 2562 นั่นคือแฮชแท็ก #ไซด์ไลน์ ขึ้นติดเทรนอันดับ 1 ดังนั้นจึงมีทวีตจำนวนมากพูดถึงเหตุการณ์นี้ตัวอย่างเช่น

“#ไซด์ไลน์ งานขายตัวแบบนี้มันทำกันเปิดเผยได้แล้วหรอว่ะ อื้มมมมม สังคมไทยตกต่ำถึงที่สุด”

“เดี๋ยวถึงกับติดเทรนได้หรอออ #ไซด์ไลน์ ”

“เศรษฐกิจไม่ดีจน #ไซด์ไลน์ ขึ้นติดเทรนทวิต”

“พยามไล่หาเหตุผลของเทรนอันดับ1 แต่หาไม่เจอ #ไซด์ไลน์ ”

คณะผู้วิจัยไม่แน่ใจสาเหตุของปรากฏการณ์นี้ แต่พบว่าทวีตที่ยกตัวอย่างทั้งหมดใช้ แฮชแท็ก #ไซด์ไลน์ ซึ่งทำให้การกรองข้อมูลด้วยแฮชแท็กยากขึ้นอีก เนื่องจากทวีตเหล่านี้ไม่ได้เกี่ยวข้องกับการค้าประเวณี เป็นเพียงการพูดถึงปรากฏการณ์ของแท็ก #ไซด์ไลน์ ติดเทรนอันดับหนึ่ง

**3.5 ความแม่นยำในการจำแนกข้อมูลแบบอัตโนมัติ**

คณะผู้วิจัยนิยามมาตรวัดประสิทธิภาพของการจำแนกข้อมูลด้วยค่าความแม่นยำ นั่นคืออัตราส่วนของ จำนวนทวีตที่เกี่ยวข้องกับการค้าประเวณี ต่อ จำนวนทวีตที่ถูกกรองมาทั้งหมด โดยมนุษย์จะเป็นผู้ตัดสินสุดท้ายว่าทวีตใดที่เกี่ยวข้องกับการการค้าประเวณี

|  |  |  |  |
| --- | --- | --- | --- |
| วิธีการ | จำนวนข้อความ  ที่ถูกจำแนก | จำนวนข้อความ  ที่จำแนกได้ถูกต้อง | ความแม่นยำ |
| จำแนกโดยคำสำคัญ | 6106 | 5758 | 94.3% |
| จำแนกโดยแฮชแท็ก | 1373 | 1349 | 98.25% |

จากการทดลองจะพบว่า การกรองโดยแฮชแท็ก มีความแม่นยำมากกว่า แต่หาข้อมูลที่เกี่ยวข้องได้น้อยกว่า หรือกล่าวได้ว่ามีความสมบูรณ์น้อยกว่า ค่าความสมบูรณ์ที่แท้จริงไม่ได้นำมาคำนวณในการทดลองนี้เนื่องจากปริมาณข้อมูลมีจำนวนมาก (4.8 ล้าน) แต่เราสามารถประมาณได้ว่า การจำแนกโดยคำสำคัญจะมีความสมบูรณ์กว่าจำแนกโดยแฮชแท็กประมาณ 4.2 เท่าจากการคำนวณ (5758/1349) ในขั้นการคำนวณหาค่าความสัมพันธ์เชิงสถิติ คณะผู้วิจัยใช้ข้อมูลจากจำแนกโดยคำสำคัญ ที่ผ่านการจำแนกโดยมนุษย์อีกขั้นหนึ่ง (5758 ทวีต) เนื่องจากมีความสมบูรณ์มากกว่าจำแนกโดยแฮชแท็ก และมีความถูกต้องเนื่องจากผ่านการจำแนกโดยมนุษย์แล้ว หากต้องการให้การจำแนกเป็นอัตโนมัติมีความแม่นยำและสมบูรณ์สูง ทางหนึ่งที่จะทำได้คือการใช้วิธีการเรียนรู้ของเครื่องจักร (Machine Learning) เข้ามาช่วย แต่หัวข้อนี้อยู่นอกขอบเขตของงานวิจัยชิ้นนี้

**4. ผลการทดลอง**

ข้อมูลจำนวนผู้ทวีตข้อความเกี่ยวกับการค้าประเวณีรายจังหวัดมีความไม่สมบูรณ์สองประการคือ คำสำคัญที่เลือกใช้มีเพียงแค่ 3 คำสำคัญเท่านั้นคือ 'ไซด์ไลน์' 'ขายตัว' และ 'ขายน้ำ' แน่นอนว่ายังคงมีคำอื่นที่จะช่วยทำให้การจำแนกสมบูรณ์มากขึ้น อีกทั้งข้อมูลนี้ยังเลือกเฉพาะทวีตที่มีพิกัดภูมิศาสตร์เท่านั้น เนื่องจากจะช่วยการระบุตำแหน่งแม่นยำขึ้นจากจำนวน 5758 ทวีต ผู้ใช้บางคนได้โฆษณาค้าประเวณีหลายครั้ง ดังนั้นจำนวนทวีตอาจจะไม่เหมาะต่อการคำนวณสถิติเพราะมีความซ้ำซ้อน ดังนั้นคณะผู้วิจัยนับจำนวนผู้ใช้ที่การโฆษณาค้าประเวณีในแต่ละจังหวัด คณะผู้วิจัยยังได้นับจำนวนผู้ใช้ที่ทวีตทั้งหมด(จาก 4.8 ล้าน ทวีต) ในแต่ละจังหวัดด้วย ในตารางด้านล่างจะเป็นตารางแสดงจำนวนผู้ใช้ที่โฆษณาเกี่ยวกับการค้าประเวณีสูงที่สุด 10 จังหวัดแรกพร้อมจำนวนผู้ใช้ทวิตเตอร์ในจังหวัดนั้น

**ตารางที่ 2**

|  |  |  |
| --- | --- | --- |
| ชื่อจังหวัด | จำนวนผู้ใช้ทวิตเตอร์ | จำนวนผู้ใช้ที่ทวีต  เกี่ยวข้องกับการค้าประเวณี |
| กรุงเทพมหานคร | 86089 | 69 |
| นครราชสีมา | 6473 | 19 |
| ระยอง | 6799 | 14 |
| ชลบุรี | 14542 | 13 |
| นนทบุรี | 10554 | 11 |
| เพชรบูรณ์ | 1659 | 11 |
| ลพบุรี | 1432 | 10 |
| สงขลา | 5177 | 8 |
| เชียงใหม่ | 11205 | 8 |
| ชุมพร | 921 | 7 |

รูปภาพด้านล่างคือ[แผนที่แถบสี (Choropleth) โดยที่สีดำจะแทนข้อมูลที่มีค่าเป็น 0 หมายถึงไม่พบผู้ใช้ที่มีการโฆษณาการค้าประเวณีระดับความเข้มของสีเขียวบอกถึงปริมาณผู้ใช้ แผนที่ด้านซ้ายแสดงจำนวนผู้ใช้ทวิตเตอร์ทั้งหมดในแต่ละจังหวัด และด้านขวาเป็น จำนวนผู้ใช้ที่โฆษณาการค้าประเวณี](https://docs.microsoft.com/th-th/power-bi/visuals/power-bi-visualization-filled-maps-choropleths)

|  |  |
| --- | --- |
|  |  |

**รูปที่ 1 จำนวนผู้ใช้ทวิตเตอร์ทั้งหมดในแต่ละจังหวัด** **รูปที่ 2 จำนวนผู้ใช้ที่โฆษณาการค้าประเวณี**

คณะผู้วิจัยได้ทดลองว่าตัวแปรทางสังคมใดมีความสัมพันธ์เชิงสถิติต่อจำนวนผู้ใช้ที่โฆษณาการค้าประเวณีรายจังหวัด โดยตัวแปรสามตัวที่คณะผู้วิจัยใช้คือ จำนวนประชากรในแต่ละจังหวัด (population) ผลิตภัณฑ์มวลรวมจังหวัด (GDP) และผลิตภัณฑ์มวลรวมจังหวัดต่อคน (GDP per capita) คณะผู้วิจัยได้คำนวณค่าสหสัมพันธ์เพียร์สัน โดยใช้ไลบรารีแพนดาส (Pandas) ค่าสหสัมพันธ์เพียร์สันของแต่ละตัวแปรเป็นตามตารางด้านล่าง

**ตารางที่ 3**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ผลิตภัณฑ์มวลรวมจังหวัด | จำนวนประชากร | ผลิตภัณฑ์มวลรวมจังหวัดต่อคน |
| จำนวนผู้ใช้ที่โฆษณาการค้าประเวณี | 0.9379 | 0.922478 | 0.452635 |

\*ข้อมูลจำนวนประชากรในแต่ละจังหวัด (population) ผลิตภัณฑ์มวลรวมจังหวัด (GDP) และผลิตภัณฑ์มวลรวมจังหวัดต่อคน (GDP per capita) นำมาจากสำนักงานสถิติแห่งชาติ โดยข้อมูลสถิตินี้เป็นข้อมูลปี พ.ศ. 2560[[4]](#footnote-4)

**5. อภิปราย**

ข้อมูลจากสำนักงานสถิติแห่งชาติ ผลิตภัณฑ์มวลรวมจังหวัด (GDP) และจำนวนประชากรของแต่ละจังหวัด (population) นั้นมีความสัมพันธ์เชิงสถิติที่สูงอยู่แล้ว โดยจากการทดลองพบค่าสหสัมพันธ์เพียร์สันเป็น 0.920916 นั่นคือจังหวัดที่มีประชากรมากก็มีแนวโน้มจะมีผลิตภัณฑ์มวลรวมจังหวัดสูง จากผลการทดลองพบว่า ผลิตภัณฑ์มวลรวมจังหวัด (GDP) และจำนวนประชากร (population) ต่างมีค่าสหสัมพันธ์เพียร์สันสูงทั้งคู่ นั่นคือ 0.9379 และ 0.922478 ตามลำดับ ดังนั้น จากตัวเลขสามารถสื่อได้ว่า ขนาดเศรษฐกิจรวมของจังหวัดมีความสัมพันธ์ต่อจำนวนการค้าประเวณีบนทวิตเตอร์ ขณะที่ผลิตภัณฑ์มวลรวมจังหวัดต่อคน (GDP per capita) มีค่า สหสัมพันธ์เพียร์สันเพียง 0.452635

กลุ่มตัวอย่างของข้อมูลชุดนี้อาจไม่สมบูรณ์จากสาเหตุที่ได้อภิปรายไปก่อนหน้านี้ นั่นคืออัลกอริธึมการสุ่มตัวอย่างของทวิตเตอร์ การเลือกเฉพาะข้อมูลที่มีพิกัดภูมิศาสตร์ และการจำแนกโดยใช้คำสำคัญ การใช้มนุษย์ในการกรองข้อมูลเพื่อลดข้อผิดพลาดของการจำแนกโดยใช้คำสำคัญอาจทำได้ในกรณีที่ข้อมูลไม่ใหญ่เกินไป (6106 ทวีต)อย่างไรก็ตาม หากข้อมูลขนาดใหญ่กว่านี้มาก การใช้มนุษย์อาจเป็นไปได้ยากในทางปฏิบัติ ดังนั้นการใช้วิธีการเรียนรู้ของเครื่องจักรจะช่วยทำให้จำแนกข้อมูลครบถ้วนสมบูรณ์กว่าเดิมได้

ภูมิศาสตร์เป็นปัจจัยสำคัญหนึ่งในการค้าประเวณี คณะวิจัยพบว่าการโฆษณาค้าประเวณีจำนวนมากจะมีการระบุพื้นที่ให้บริการ เช่น

“#รับงานหัวหิน #รับงานปราณบุรี #รับงานชะอำ #slหัวหิน #slปราณบุรี #slชะอำ”

“#รับงานไซด์ไลน์ #รับงานกทม #สาวสองรับงานนนทบุรี #สาวสองมีงู #สาวสองรับงาน”

“ผู้ชายขายน้ำ ระยองทักได้นะ 1น้ำ1000 2น้ำ1500”

“#รับงาน #ไซด์ไลน์ มหาชัยเท่านั้น ไม่เดินสายนะคะ รับโรงแรมบาหลีการ์เด้นอินน์ “

จะเห็นว่าตัวอย่างทั้งสี่ทวีตด้านบนมีการระบุสถานที่ให้บริการทั้งหมด ข้อมูลในรูปที่ 1 และ 2 ใช้ข้อมูลเพียงสถิติของจังหวัดที่โพสทวีตเท่านั้น แต่ไม่ได้ครอบคลุมพื้นที่บริการ เนื่องจากความลำบากด้านเทคนิคในการประมวลผลภาษาธรรมชาติดังที่ได้กล่าวในหัวข้อการจำแนกจังหวัดของทวีต ตัวอย่างพิกัดที่โพสทวีตและพื้นที่ให้บริการเป็นคนละจังหวัด

“#รับงานนครปฐม #รับงานอ้อมน้อย #รับงานอ้อมใหญ่ #รับงานสาย4”

ทวีตด้านบนพิกัดอยู่ที่กรุงเทพมหานคร ข้อมูลชื่อสถานที่ในทวีต ระบุเป็น “บางแคเหนือ, ประเทศไทย” แต่พื้นที่ให้บริการครอบคลุมจังหวัดนครปฐมดังแสดงในข้อความ ในกรณีนี้สถิติจะนับเป็นเพียงแค่จังหวัดกรุงเทพมหานครเท่านั้น สถิติที่เกี่ยวข้องกับพื้นที่ให้บริการที่ระบุอยู่ในข้อความนั้นจำเป็นต้องใช้เทคนิคประมวลผลด้านภาษาเข้ามาช่วยเพื่อประมวลผลข้อมูลขนาดใหญ่นี้ซึ่งอยู่นอกเหนือขอบเขตของงานวิจัยชิ้นนี้

วิธีการติดต่อและราคาเป็นอีกข้อมูลที่ผู้ให้บริการมักจะให้ข้อมูลไว้ในทวีต โดยวิธีการติดต่อมักจะใช้เป็นเบอร์โทรศัพ์และไลน์ ตัวอย่างเช่น

“#สาวรับงานบ้านฉาง ไลน์ xxxx ค่ะ โทร xxxxxx ค่ะ #สาวไซด์ไลน์บ้านฉาง”

“#รับงานไซด์ไลน์. ค่าตัว 2000.ต่อ1ช.ม ไอดีไลน xxxxx. #ปู่เจ้า #รับงานสมุทปราการ #เทพารักษ์”

ผู้ให้บริการอาจจะระบุ เพศและลักษณะทางกายภาพ เช่น

“#ขายตัว สาวสองทำนมรับงานอโศกมีห้องค่ะอายุ xx น้ำหนักxx สูงxxนมxx ผิวขาว”

คณะผู้วิจัยคำนวณผู้ให้บริการที่ใช้คำว่าสาวสองในทวีตพบว่ามีจำนวนประมาณ 10% ดังตารางด้านล่าง

**ตารางที่ 4**

|  |  |
| --- | --- |
|  | จำนวนผู้ใช้ |
| ผู้ให้บริการทั้งหมด | 564 |
| ผู้ให้บริการที่ระบุคำว่าสาวสองในทวีต | 58 |

การกลั่นแกล้งหรือการหลอกลวงต่อผู้ให้ผู้บริการนั้น คณะผู้วิจัยคาดว่ามีอยู่ อย่างไรก็ตามการเก็บข้อมูลส่วนนี้ผ่านทวีตอาจเป็นเรื่องยาก การศึกษาเรื่องนี้แบบสอบถามเชิงลึกอาจเป็นตัวเลือกที่เหมาะสมกว่า ตัวอย่างของทวีตที่สามารถอนุมานว่ามีการหลอกลวงหรือกลั่นแกล้งให้ผู้บริการ

“รับงาน เองคะ แอดไลเลยนะคะ ขอพี่ๆที่นัดจริงไม่เอาหลอกนะคะ”

“ขอคนนัดจริงนะคะ เกรียนคีย์บอร์ด อย่าทักมา”

เหตุผลที่เข้ามาสู่อาชีพค้าประเวณีบนสื่อสังคมออนไลน์นั้นจำเป็นต้องแบบสอบถามเชิงลึกเช่นกันอย่างไรก็ตาม ปัจจัยด้านเศรษฐกิจน่าจะเป็นเหตุผลหลักเช่นเดียวกับการค้าประเวณีแบบดั้งเดิมเช่นการค้าประเวณีแฝงใน สถานบริการนวดแผนโบราณ (ภูวาเดช โหราเรือง, 2018) ตัวอย่างของทวีตที่สามารถอนุมานถึงปัจจัยด้านเศรฐกิจมีดังต่อไปนี้

“ว่างทั้งวัน #รับงานพระพุทธบาท #ไซด์ไลน์ #ขายตัว #ร้อนเงิน @line:xxxx”

“ร้อนเงินครับทักมาคุยครับ#ขายตัวเชียงราย”

คณะผู้วิจัยพบว่าภูมิศาสตร์มีผลต่อกิจกรรมการค้าประเวณีบนสื่อสังคมออนไลน์ โดยจังหวัดที่มีจำนวนประชากรและผลิตภัณฑ์มวลรวมจังหวัดสูง จะมีแนวโน้มการค้าประเวณีบนสื่อสังคมออนไลน์มากขึ้น เศรษฐกิจเป็นปัจจัยสำคัญต่อกิจกรรมการค้าประเวณีบนสื่อสังคมออนไลน์ อย่างไรก็ตาม คณะผู้วิจัยพบว่าขนาดเศรษฐกิจรวมมีความสัมพันธ์เชิงสถิติมากกว่าต่อคน หากทำให้การค้าประเวณีถูกกฎหมายในประเทศไทย นักวิจัยคงสามารถเข้าถึงข้อมูลทางสถิติที่แม่นยำมากขึ้นโดยการเก็บข้อมูลแบบทางการจากภาครัฐและหน่วยงานที่เกี่ยวข้อง

**เอกสารอ้างอิง**

1. กนิษฐา ไชยแสง, & ขจร ฝ้ายเทศ. (2014). การใช้ทวิตเตอร์ของผู้สื่อข่าวโทรทัศน์ในการนำเสนอข้อมูล ข่าวสาร.
2. พิมพ์วิภา มีงามดี, สาธนี นาแซง, & พิชญาวี ทองกลาง. (2020). การนำเสนอภาพลักษณ์ของ “เกย์” และ “เลสเบี้ยน” ใน ทวิตเตอร์. วารสาร วิชาการ คณะ มนุษยศาสตร์ และ สังคมศาสตร์ มหาวิทยาลัยราชภัฏ นครราชสีมา, 1(1), 30-30.
3. ภูวาเดช โหรา เรือง. (2018). เศรษฐกิจใต้ดิน: การ ค้าประเวณี แฝง ใน สถาน บริการ นวด แผน โบราณ ใน ประเทศไทย. Journal of Rangsit Graduate Studies in Business and Social Sciences, 4(2), 201-211.
4. อารยา พุดตาล. การวิเคราะห์ทวิตเตอร์ในเหตุการณ์สวรรคตของพระบาทสมเด็จพระปรมินทรมหาภูมิพล อดุลยเดช (Doctoral dissertation, จุฬาลงกรณ์ มหาวิทยาลัย).
5. Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? Geolocation and language identification in Twitter. The Professional Geographer, 66(4), 568-578.
6. Pfeffer, J., Mayer, K., & Morstatter, F. (2018). Tampering with twitter’s sample api. EPJ Data Science, 7(1), 50.
7. Vateekul, P., & Koomsubha, T. (2016, July). A study of sentiment analysis using deep learning techniques on Thai Twitter data. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 1-6). IEEE.

1. https://www.tweepy.org/ [↑](#footnote-ref-1)
2. <https://docs.python.org/3/library/pickle.html> [↑](#footnote-ref-2)
3. <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object> [↑](#footnote-ref-3)
4. http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector\_10\_28\_TH\_.xls [↑](#footnote-ref-4)