



## CPE 213 Data Models

### Proposal

## Hotel Recommendation in Europe

### Group member

นางสาวกัญญารักษ์	สดโพธิ์	63070501003
นายพิทยา	คงธนสถิตย์	63070501051
นางสาวมณฑิตา	อานุภาพสโมสร	63070501056
นางสาวศรัลวรลิณ	ว่องอุดมชนกุล	63070501058
นายศิลา	ศิลาภีรัตพงศ์	63070501060

### Present

Dr. Santitham Prom-on

King Mongkut's University of Technology Thonburi

Computer Engineering, Faculty of Engineering

Semester 2 / 2021

## Abstract

ในปัจจุบันนี้การไปท่องเที่ยวในแต่ละครั้งการเลือกเข้าพักโรงแรมในแต่ละโรงแรมเป็นสิ่งที่จำเป็นสำหรับการท่องเที่ยวซึ่งมีหลายปัจจัยที่ผู้เข้าพักจะใช้ในการตัดสินใจมีการเข้าพักโรงแรมนั้นๆไม่ว่าจะเป็นรีวิวของผู้ที่เคยเข้าพักโรงแรมนั้นด้านบวกหรือด้านลบ การให้คะแนนจากผู้เข้าพักของแต่ละโรงแรมซึ่งผู้เข้าพักอาจจะต้องการคำแนะนำ หรือการจัดอันดับโรงแรมที่ดีที่สุดและเหมาะสมที่สุดสำหรับตัวนักท่องเที่ยว เราจึงนำข้อมูลเรื่อง Hotel Reviews Data in Europe ซึ่งข้อมูลที่น่าสนใจวิเคราะห์เบื้องต้น ได้แก่ ข้อความเชิงบวก ข้อความเชิงลบ คะแนนรีวิว สัญชาติของผู้รีวิว และ ที่อยู่ของโรงแรม ในระหว่างปี 2015 - 2017เพื่อนำไปใช้ประโยชน์ในการแนะนำโรงแรมให้แก่ผู้ที่กำลังหาโรงแรม เพื่อหาโรงแรมที่คาดว่าจะตรงกับความสนใจของผู้ที่กำลังหาโรงแรมเข้าพัก โดยโมเดล Recommendation system แบบเลือกตัวอย่างโรงแรมมา 1 แห่ง (แทนโรงแรมแรกที่ผู้เข้าพักได้เลือกกดดู) แล้วหาค่าความคล้าย (Cosine Similarity) กับโรงแรมอื่น โดยการใช้ทฤษฎีการทำระบบ Recommendation System (Content-Based Filtering (CBF) Algorithm)

## Introduction

ในทวีปยุโรปมีโรงแรมให้เลือกเข้าพักหลายแห่งทำให้การตัดสินใจเลือกเข้าพักในโรงแรมแต่ละที่ตามความต้องการเป็นไปได้ยาก สิ่งหนึ่งที่สามารถช่วยในการตัดสินใจได้ คือ รายละเอียดการรีวิวในปี 2015 ถึง 2017 โดยเราจะนำข้อมูลข้อความเชิงบวก (Positive\_review) , ข้อความเชิงลบ (Negative\_review) , คะแนนรีวิว (Average\_rating) , สัญชาติของผู้รีวิว (Reviewer\_Nationality) และ ที่อยู่ของโรงแรม (Hotel\_Address) นำมาวิเคราะห์เบื้องต้นก่อนนำไปใช้ประโยชน์ในการแนะนำโรงแรมให้แก่ผู้ที่กำลังหาโรงแรมเข้าพัก

ทั้งนี้เราจึงได้เลือก Recommendation System มาแนะนำโรงแรมให้แก่ผู้ที่กำลังหาโรงแรมเข้าพัก โดยเราจะทำการหาค่าความคล้าย (Cosine Similarity) จากกลุ่มโรงแรมตัวอย่าง เพื่อหาโรงแรมที่คาดว่าจะตรงกับความสนใจของผู้ที่กำลังหาโรงแรมเข้าพัก ซึ่งทำให้ง่ายต่อการตัดสินใจในการเลือกเข้าพักโรงแรมให้ตรงกับวัตถุประสงค์ที่ตามต้องการ

## Data Description

ข้อมูลนี้แสดงรายละเอียดการรีวิวโรงแรมของลูกค้ากว่า 515000 รีวิวซึ่งเป็นรีวิวของโรงแรมในทวีปยุโรปจำนวน 1493 โรงแรม ประกอบไปด้วย

- `Hotel_Address`: ที่อยู่ของโรงแรม
- `Review_Date`: วันที่ผู้เขียนรีวิวโพสต์รีวิวที่เกี่ยวข้อง
- `Hotel_Name`: ชื่อของโรงแรม
- `Reviewer_Nationality`: สัญชาติของผู้เขียนรีวิว
- `Negative_Review`: ความคิดเห็นเชิงลบที่มีต่อโรงแรม  
(ถ้าหากผู้เขียนรีวิวไม่ได้แสดงความคิดเห็นเชิงลบ จะแสดงผล 'No Negative')
- `ReviewTotalNegativeWordCounts`: จำนวนคำทั้งหมดในความคิดเห็นเชิงลบ
- `Positive_Review`: ความคิดเห็นเชิงบวก ที่ผู้เขียนรีวิวให้กับโรงแรมนั้นๆ  
ถ้าผู้เขียนรีวิวไม่ได้แสดงความคิดเห็นเชิงบวกไว้ จะแสดงผลเป็น 'NoPositive'
- `ReviewTotalPositiveWordCounts`: จำนวนคำทั้งหมดในความคิดเห็นเชิงบวก
- `Reviewer_Score`: คะแนนที่ผู้เขียนรีวิว ให้กับโรงแรมนั้นๆ จากประสบการณ์เข้าเยี่ยมชมของผู้เขียนรีวิว
- `TotalNumberOfReviewsReviewerHasGiven`: จำนวนรีวิวที่ผู้เขียนรีวิวเคยให้ไว้ในอดีต
- `TotalNumberOf_Reviews`: จำนวนรีวิวที่ถูกต้องทั้งหมดที่โรงแรมมี
- `Tags`: แท็กผู้ที่รีวิวให้โรงแรม
- `daysincereview`: ระยะเวลาระหว่าง review date และ scrape date.
- `AdditionalNumberOf_Scoring`: ยังมีแขกบางคนที่ไม่ให้คะแนนบริการมากกว่ารีวิว  
ตัวเลขนี้ระบุจำนวนคะแนนที่ถูกต้องโดยไม่ต้องตรวจสอบ
- `lat`: ละติจูดของโรงแรม
- `lng`: ลองจิจูดของโรงแรม

## Problem statement

จากปัญหาการหาโรงแรมที่ตรงกับความต้องการของผู้เข้าพักนั้นเป็นเรื่องที่ยาก ทางเราจึงได้จัดทำโมเดลเพื่อที่จะแนะนำโรงแรมที่ตรงกับความต้องการของผู้ที่กำลังหาโรงแรมเข้าพัก โดยการวิเคราะห์ข้อมูลเบื้องต้นจากข้อมูล ดังนี้ ข้อความเชิงบวก (Positive\_review) , ข้อความเชิงลบ (Negative\_review) , คะแนนรีวิว (Average\_rating) , สัญชาติของผู้รีวิว (Reviewer\_Nationality) และ ที่อยู่ของโรงแรม (Hotel\_Address)

## Proposed analytic technique

### ขั้นตอนที่ 1 คือการสำรวจข้อมูล (Data Exploration)

เพื่อวิเคราะห์ข้อมูลเบื้องต้นก่อนนำไปใช้ประโยชน์ในขั้นตอนการทำ Recommendation system โดยเลือกปัจจัยที่อาจส่งผลต่อการเข้าพักโรงแรม ได้แก่ ข้อความเชิงบวก ข้อความเชิงลบ คะแนนรีวิว สัญชาติและที่อยู่ของโรงแรมจากผู้รีวิวก่อนหน้าของแต่ละโรงแรม

- ข้อความเชิงบวก (Positive\_review) โดยการทำ Word Cloud จากการสุ่มเลือกโรงแรมมาแสดงผล และวิเคราะห์ข้อมูลเบื้องต้น
- ข้อความเชิงลบ (Negative\_review) โดยการทำ Word Cloud จากการสุ่มเลือกโรงแรมมาแสดงผล และวิเคราะห์ข้อมูลเบื้องต้น
- คะแนนรีวิว (Average\_rating) โดยคำนวณมาจากคะแนนที่ผู้เขียนรีวิวให้ (Reviewer\_Score) โดยทำการ group โรงแรมก่อน แล้วทำการหาค่าเฉลี่ยของคะแนนที่ผู้รีวิวให้
- สัญชาติของผู้เขียนรีวิว (Reviewer\_Nationality) จากการสุ่มเลือกโรงแรมมาแสดงผลและวิเคราะห์ข้อมูล โดยทำการ group สัญชาติของผู้รีวิวก่อน แล้วสร้าง bar chart พร้อมกับหาฐานนิยม
- ที่อยู่ของโรงแรม (Hotel\_Address) โดยจะทำการ group โรงแรมก่อน แล้วทำการแยกส่วนในข้อมูลที่อยู่ที่เป็นประเทศออกมา แล้วสร้างเป็น bar chart พร้อมกับหาฐานนิยม

### ขั้นตอนที่ 2 คือ การทำ recommendation system

เพื่อแนะนำโรงแรมที่เหมาะสมให้กับผู้เข้าพักจะทำการเลือกหาโรงแรมที่มีคุณสมบัติใกล้เคียงกับโรงแรมแรก que ผู้เข้าพักได้เลือกกดดู โดยการสร้าง recommendation system แบบเลือกตัวอย่างโรงแรมมา 1 แห่ง (แทนโรงแรมแรก que ผู้เข้าพักได้เลือกกดดู) แล้วหาค่าความคล้าย (Cosine Similarity) กับโรงแรมอื่น

โดยการใช้ทฤษฎีการทำระบบ Recommendation System (Content-Based Filtering (CBF) Algorithm)

- เริ่มจากการ หาค่าความเหมือนของข้อความเชิงบวก (Positive review) ของโรงแรม A เทียบกับข้อความเชิงบวก (Positive review) ของโรงแรม B ได้จากการเอา feature vector ของแต่ละข้อความเชิงบวก (Positive review) ของโรงแรมแต่ละโรงแรม มาเปรียบเทียบกับ โดยการหา Cosine Similarity โดยค่าที่ได้ออกมาจะมีค่าตั้งแต่ 0 - 1 ซึ่งค่า 0 หมายถึงไม่เหมือนเลย และค่า 1 คือเหมือนมาก

$$\mathbf{A} = \begin{bmatrix} Cat_1 \\ Cat_2 \\ \vdots \\ Cat_m \\ Tag_1 \\ Tag_2 \\ \vdots \\ Tag_n \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$