



# ระบบแนะนำภาพยนตร์ (Movie Recommendation System)

ภคชฎีณที่ ผณินทรารักษั, ศรารุจ วรคัณวิวัฒน์กุล และ ผศ.ดร.วรารณณ์ วิยานนท์  
หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

## Abstract

Movie recommendation system plays an important role in helping users to find the relevant movies from many existing movies which would reduce the time of searching for movies. This project used the datasets from Kaggle consisting of movie metadata and their rating. To find the weighted rating of each movie we did Demographic Filtering using IMDB's weighted rating formula. The formula is calculated from vote count and vote average of a particular film, the average of all movie votes and the defined minimum number of votes. In addition, we used collaborative filtering for recommending movies to users based on their similarity using user similarity and item similarity. If the weighted rating of a movie is high, it shows that the movie has high popularity which can be used to recommend it to general users. The result from Collaborative Filtering using item similarity is used to recommend similar movies to a specific user in order to decrease the time for searching movies. As for the Collaborative Filtering in User Similarity, the movie recommendation system for individual users will be determined based on similarities with other users.

## องค์ความรู้ที่เกี่ยวข้อง

- 1. Demographic filtering [1] ใช้เพื่อเสนอคำแนะนำเกี่ยวกับภาพยนตร์สำหรับผู้ทั่วไป โดยพิจารณาจากความนิยมของภาพยนตร์มีโอกาสสูงที่ผู้ใช้จะชอบ สูตรการคำนวณอันดับของภาพยนตร์แบบถ่วงน้ำหนักดังสมการ(1)  $weighted\ rating\ (WR) = \frac{v}{v+m} \times R + \frac{m}{v+m} \times C$  สมการ(1) IMDB weighted rating formula R = ค่าเฉลี่ยโหวตของภาพยนตร์แต่ละเรื่อง (vote\_average) v = จำนวนโหวตของภาพยนตร์ (vote\_count) m = จำนวนโหวตขั้นต่ำที่ต้องมีของภาพยนตร์ C = ค่าเฉลี่ยโหวตของภาพยนตร์ทั้งหมด
- 2. Collaborative filtering User Similarity [2] เป็นการแนะนำโดยหาคุณค่าที่มีพฤติกรรมเหมือนกัน สูตรการหาความคล้ายกันระหว่าง 2 เวกเตอร์ดังสมการ(2)  $Cosine\ Similarity = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum AB}{\sqrt{\sum A^2} \sqrt{\sum B^2}}$  สมการ(2) Cosine Similarity A = การให้คะแนนภาพยนตร์ของผู้ใช้ A B = การให้คะแนนภาพยนตร์ของผู้ใช้ B Item Similarity [3] เป็นการแนะนำโดยหาภาพยนตร์ที่มีความคล้ายคลึงกับที่ลูกค้าต้องการโดยเราจะคำนวณหาความเหมือนกันของไอเทมเพื่อนำมาแนะนำให้แก่ลูกค้า สูตรการหาทิศทางความสัมพันธ์ระหว่างสองตัวแปรดังสมการ(3)  $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$  สมการ(3) Correlation [4] r = ค่าสหสัมพันธ์ x = ค่าการให้คะแนนภาพยนตร์ x y = ค่าการให้คะแนนภาพยนตร์ y

## ผลการดำเนินงาน

- 1. Demographic filtering นำข้อมูลของภาพยนตร์(movie\_metadata.csv) มาคำนวณหาอันดับความชื่นชอบของผู้ใช้แบบถ่วงน้ำหนักโดยใช้ IMDB weighted rating จากสมการ(1) เพื่อนำไปแนะนำภาพยนตร์ให้กับผู้ใช้โดยภาพรวมดังแสดงในภาพที่ 2

title	vote_count	vote_average	score
The Shawshank Redemption	8358.0	8.5	8.445869
The Godfather	6024.0	8.5	8.425439
Dilwale Dulhania Le Jayenge	661.0	9.1	8.421453
The Dark Knight	12269.0	8.3	8.265477
Fight Club	9678.0	8.3	8.256385
Pulp Fiction	8670.0	8.3	8.251406
Schindler's List	4436.0	8.3	8.206639
Whiplash	4376.0	8.3	8.205404
Spirited Away	3968.0	8.3	8.196055
Life Is Beautiful	3643.0	8.3	8.187171

ภาพที่ 2 อันดับความชื่นชอบของผู้ใช้แบบถ่วงน้ำหนัก

- 2. Collaborative Filtering User Similarity นำข้อมูลของภาพยนตร์(movie\_metadata.csv) และข้อมูลความชื่นชอบของผู้ใช้(rating.csv) มาคำนวณโดยใช้สูตร Cosine Similarity จากสมการ(2) โดยหาลำดับผู้ใช้จำนวน 5 คน ที่มีความคล้ายกับผู้ใช้ที่มีลำดับที่ 1

[1, 120, 81, 49, 125, 91]

	1	120	81	49	125	91
1	1.000000	0.529576	0.494071	0.480652	0.476662	0.463316
2	0.124295	0.178544	0.223689	0.124236	0.140154	0.186631
3	0.118821	0.137621	0.136526	0.151552	0.175610	0.137960
4	0.103646	0.100012	0.154947	0.117456	0.164447	0.111659

ภาพที่ 3 ผู้ใช้ที่มีความคล้ายกับผู้ใช้ลำดับที่ 1

ภาพที่ 4 แสดงค่าความคล้ายของผู้ใช้ลำดับที่ 1

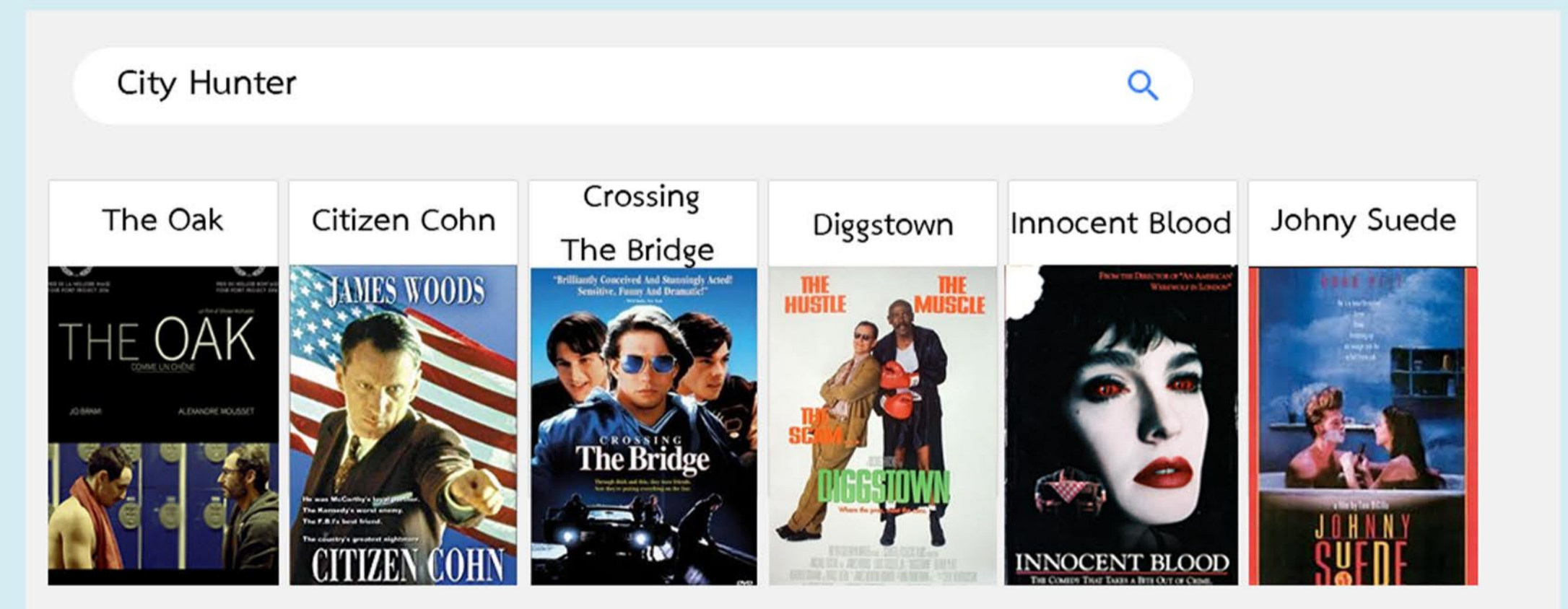
- Item Similarity นำข้อมูลของภาพยนตร์(movie\_metadata.csv) ข้อมูลความชื่นชอบของผู้ใช้(rating.csv) และข้อมูลเลขภาพยนตร์(link.csv) มาคำนวณหาความสัมพันธ์(Correlation) จากสมการ(2)

Title	City Hunter	Protocol	Visions of Light	Shipwrecked	The Fourth Man	Vabank
City Hunter	1.000000	0.929341	0.865292	0.862026	0.858710	0.822185
The Oak	-0.002424	-0.002109	-0.001811	-0.002049	-0.001757	-0.002748
Citizen Cohn	0.322899	-0.002109	0.241172	0.513554	0.194708	-0.002748
Crossing The Bridge	0.210962	-0.002978	0.157567	0.336885	0.126988	-0.003861
Diggstown	-0.002424	-0.002109	-0.001811	-0.002049	-0.001757	-0.002748
Innocent Blood	0.322899	-0.002109	0.241172	0.513554	0.194708	-0.002748
Johnny Suede	-0.002424	-0.002109	-0.001811	-0.002049	-0.001757	-0.002748

movieid	title	similarity	
9033	26765	City Hunter	1.000000
7427	7727	Protocol	0.929341
9050	26797	Visions of Light	0.865292
8979	26690	Shipwrecked	0.862026
8871	26493	The Fourth Man	0.858710
5800	5912	Vabank	0.822185

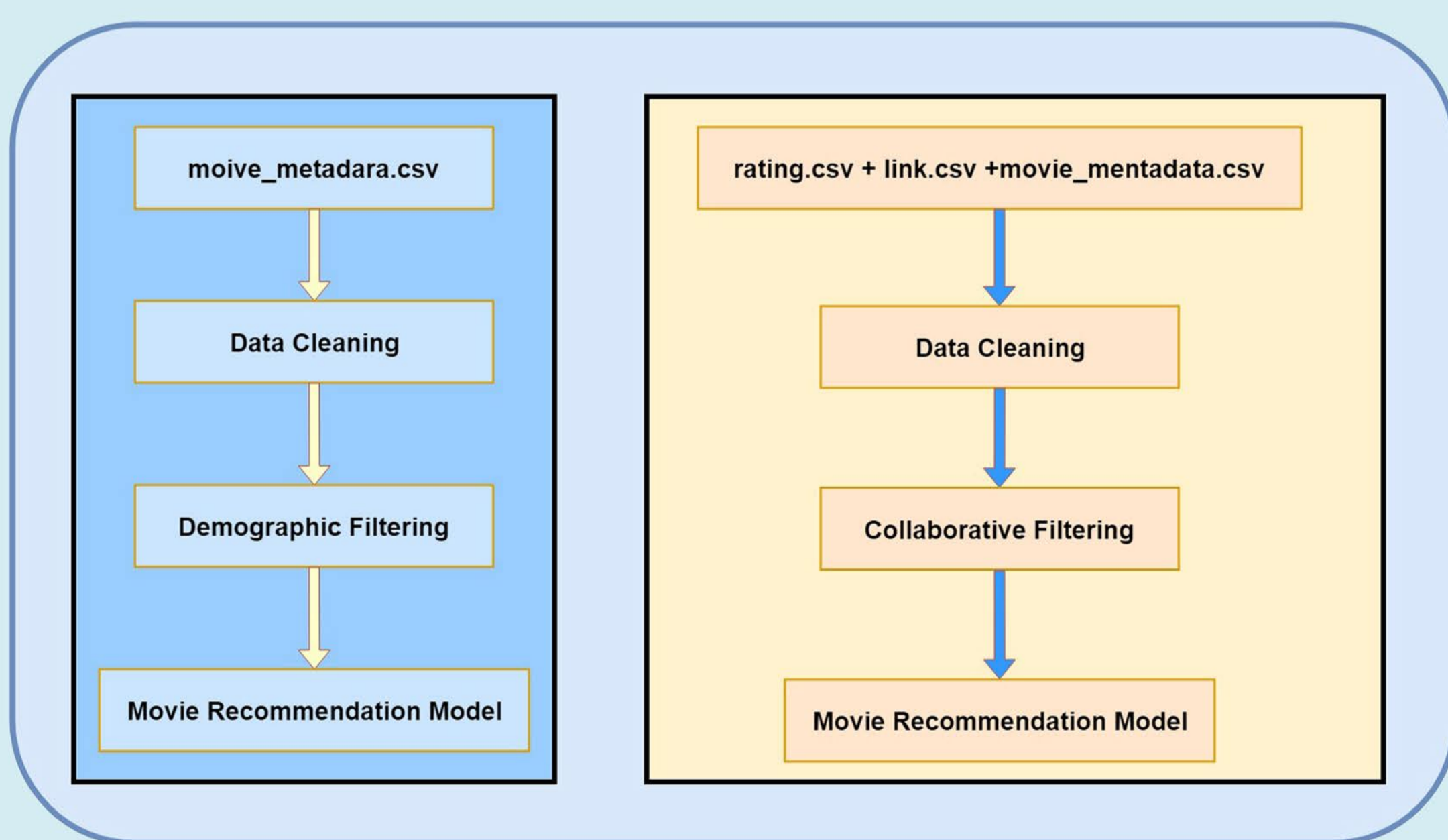
ภาพที่ 5 แสดงผลลัพธ์ของการคำนวณสูตร Correlation

ภาพที่ 6 แสดงผลลัพธ์ภาพยนตร์ที่คล้ายกับ City Hunter



ภาพที่ 7 แสดงการค้นหาภาพยนตร์เรื่อง City Hunter

## ขั้นตอนการดำเนินงาน



ภาพที่ 1 ภาพรวมของการดำเนินงาน

ชุดข้อมูลที่ใช้มาจากเว็บไซต์ Kaggle ใช้ชุดข้อมูลของ The Movie Dase โดยจะแสดงเพียงข้อมูลบางคอลัมน์ที่สำคัญ ดังตารางที่ 1, 2 และ 3

ตารางที่ 1 ข้อมูลภายในไฟล์ rating.csv

ชื่อคอลัมน์	รายละเอียดของคอลัมน์
userid	รหัสของผู้ใช้
movieid	รหัสของภาพยนตร์
rating	คะแนนของภาพยนตร์

ตารางที่ 2 ข้อมูลภายในไฟล์ link.csv

ชื่อคอลัมน์	รายละเอียดของคอลัมน์
movieid	รหัสของภาพยนตร์
tmdbid	รหัสภาพยนตร์ของ The movie database

ตารางที่ 3 ข้อมูลภายในไฟล์ movie\_metadata.csv

ชื่อคอลัมน์	รายละเอียดของคอลัมน์
id	รหัสภาพยนตร์ของ imdb
title	ชื่อภาพยนตร์อย่างเป็นทางการ
vote_average	คะแนนเฉลี่ยของภาพยนตร์
vote_count	จำนวนคนโหวต

## สรุปผลการดำเนินงาน

จากหลักการ Demographic Filtering ทำให้ได้อันดับภาพยนตร์แบบถ่วงน้ำหนักของภาพยนตร์แต่ละเรื่อง ถ้ามีค่ามากแสดงให้เห็นถึงภาพยนตร์เรื่องนั้นมีความนิยมสูง เพื่อนำไปแนะนำภาพยนตร์ให้กับผู้ใช้โดยภาพรวม และจากหลักการ Collaborative Filtering แบบ Item Similarity ได้ค่าความคล้ายกันของแต่ละภาพยนตร์ เพื่อเป็นการแนะนำผู้เฉพาะบุคคลในการค้นหาภาพยนตร์ ส่วน Collaborative Filtering แบบ User Similarity จะได้รับระบบแนะนำภาพยนตร์สำหรับผู้ใช้แต่ละคนโดยหาจากค่าความคล้ายคลึงกับผู้ใช้อื่นและทำการพัฒนาแอปพลิเคชันต้นแบบสำหรับการแนะนำภาพยนตร์ต่อไป

## บรรณานุกรม

- [1] Dooms, S., Bellogin, A., Pessemier, T. and Martens, L. (2016). A Framework for Dataset Benchmarking and Its Application to a New Movie Rating Dataset. ACM Transactions on Intelligent Systems and Technology, 7(3), pp.1-28.
- [2] Phantunin, M. and Chirawichitchai, A. (2017). A PERSONALIZED RECOMMENDER SYSTEM USING USER-BASED AND ITEM-BASED COLLABORATIVE FILTERING METHOD. [online] Available at: https://www.east.spu.ac.th/journal/booksearch/upload/1682-Ratawan.pdf [Accessed 27 Mar. 2019].
- [3] Eresearch.library.ssu.ac.th. (2019). [online] Available at: http://eresearch.library.ssu.ac.th/bitstream/123456789/252/6/ird\_083\_55%20%285%29.pdf [Accessed 29 Mar. 2019].
- [4] Medium. (2019). สถิติเบื้องต้นง่ายๆ ที่จะทำให้คุณเข้าใจการวิเคราะห์มากขึ้น (ตอนที่ 2). [online] [Accessed 27 Mar. 2019].