**Market Basket Analysis by Association Rule Approach**

**using Various of Machine Learning Algorithm**

**Introduction:**

In the early time of modern trading of mankind , most of the trading activity will focus on the process of how to maximize the capacity of supply chain in order to produce the highest quantity of product and deliver to the client in the shortest period of time. This period is known as the “product Era” and “Sales Era”. After a period of time, the trends of the world tend to shift towards the trends which pay more attention on the customer need and this era is known as the “Marketing Era”. This trend of focusing on the need of customer has created the marketing strategy known as CRM (Customer Relationship Management). CRM is the strategy that not only connect the seller with the customer, but it can also be considered as a tool that capable of analyze the behavior of the customer for further business marketing development.

ในช่วงเริ่มแรกของการค้าขายนั้น ตลาดทั่วไปจะตั้งเป้าไปที่การผลิตสินค้าออกมาเป็นจำนวนมาก และจำหนายสินค้าออกไปให้ได้อย่างรวดเร็ว เราเรียกยุคเหล่านี้ว่า Product Era และ Sales Era . เมื่อเวลาผ่านไปก็ได้กำเนิดเป็น marketing era ที่เริ่มมีการกล่าวถึงกลยุทธ์ในการขาย แล้วจึงเกิดการตระหนักถึงลูกค้าที่เป็นอีกปัจจัยในการทำการขาย จึงเกิดขึ้นมาเป็นกลยุทธ์ที่ชื่อว่า CRM ในที่สุด. CRM ไม่ได้เป็นเพียงแต่การติดต่อกับลูกค้าเพียงเท่านั้น แต่รวมถึงการวิเคราะห์พฤติกรรมของพวกเขาเพื่อที่จะนำไปวางแผนการตลาดต่อไป

In the present period of time there are countless number of the software to help with the marketing solution for the business which cover all area of marketing area, one area that is significantly important is the customer relationship management and the example of the system which is available for this task is the software for customer relationship and data analytic. This software has pay an important role for the business by helping in the process of collecting the business data which include the sell, shipping and customer preference data and analyze for better business strategy in a the real time manner. Moreover it has also pay important role in the process of minimizing the expenses and mistakes coursing by the company’s human resource too.

โดยในปัจจุบัน เทคโนโลยีต่างๆ เช่น ซอฟต์แวร์ลูกค้าสัมพันธ์ หรือ ระบบ data analytic ได้เข้ามามีส่วนช่วยในการอำนวยความสะดวกในเรื่องของการเก็บและวิเคราะห์ข้อมูล ทำให้การวางแผน strategy สำหรับองค์กรหรือผู้ประกอบการเป็นไปได้อย่างง่าย ทันการณ์ (realtime) อีกทั้งยังช่วยลดค่าใช้จ่ายและข้อผิดพลาดในส่วนของทรัพยากรบุคคลลงอีกด้วย.

The data analytic system is the system that is capable of study the behavior of the data input by the process of recognize the pattern created by that data for the purpose of predicting the next action that might happen accordingly, in the marketing area the data is the behavior of the human that are collected and the system will predict the action that is most likely to be performed. In February 2021 the scientist from south Korea name Seunghyun Choi and the fellow scientist has published an article which mention about the application of sequential pattern strategy in order to classify the event that could have taken place from the monitoring process of human being in the limited area and predict the most possible upcoming event by the help of Object detection algorithm. Wherewith the algorithm will simulate the pattern of event that could have happen accordingly which related to the data selected.

data analytic นั้นมีความสามารถในการศึกษาพฤษติกรรมของมนุษย์จากการจดจำ pattern ต่างๆ แล้วทำนาย action ถัดไปที่อาจเกิดขึ้นในอนาคตได้. เมื่อเดือนกุมภาพันธ์ ปี 2021 ที่ผ่านมา นักพัฒนาชาวเกาหลีใต้ seunghyun choi และทีมได้ตีพิมพ์บทความที่กล่าวถึงการใช้ sequential pattern strategy สำหรับ classify เหตุการณ์ต่างๆจากการ monitor พฤติกรรมของมนุษย์ในพื้นที่จำกัด และทำนายเหตุหารณ์ที่อาจจะเกิดขึ้นจาก object detection algorithm โดย algoritm สามารถจำลอง pattern ของเหตุการณ์ที่อาจเกิดขึ้นต่อจากกันซึ่งอ้างอิงจากข้อมูลที่รวบรวมได้.

There are many studies related to the technology that apply the knowledge about customer behavior prediction, for example in 2013 scientist name Laura Badea from Bucharest University has applied Artificial Neural Networks (ANN) in the study in order to detect the character of individual which is highly possible to have chance to deposit money from the information which includes race, gender, educational background, age and wage. And the result of the algorithm came out were considered to be positively promising with 73 to 78 percentage of accuracy rate of prediction. Even though this algorithm were able to perform the task of prediction in the rate that could considered to be in a good performance, but this technology contains some downturn that has to be considerate too which include the time consuming and or the higher chance that could leads to the overfitting scenario compare to the other method available.

ยังมีอีกหลายการศึกษาที่กล่าวถึงการใช้เทคโนโลยีในการประยุคต์ใช้กับการทำนายพฤติกรรมของลูกค้า เช่น ในปี 2013 Laura Badea จาก Bucharest University ได้ใช้ Artificial Neural Networks (ANN) สำหรับตรวจสอบลักษณะของปัจเจคบุคคลที่มีโอกาสที่จำทำรายการฝากเงินจากข้อมูลส่วนบุคคล เช่น เชื้อชาติ เพศ การศึกษา อายุ หรือ รายได้ โดยผลลัพธ์แสดงให้เห็นว่าอัลกอริทึมสามารถตรวจจับกลุ่มตัวอย่างที่มีโอกาสฝากและไม่ฝากมีความแม่นยำของโมเดลอยู่ระหว่าง 73-78 % แม้อัตราการตัวจับจะอยู่ในระดับสูง แต่เทคโนโลยีดังกล่าวก็ยังมีเรื่อง่ที่ต้องนำมาพิจารณาเพิ่มเติมเช่น time consuming หรือ chance ที่จะเกิดสถารการณ์ overfitting ที่เพิ่มขึ้นเมื่อเปรียบเทียบกับ algorithm อีกชนิดหนึ่ง

One of the most simple method that is available nowadays that is wildly applied in the CRM technique in order to identify the relationship between each type of product that have high possibility to be purchased analyzed from the products that is already put in the basket is called market basket analytic. This technique is process with the help of the association rule algorithm. This technique is considered to be an efficient tool for the organization for analyzing the purchasing behavior of the customer for further marketing strategy. Market basket analytic can be applied via various tools and countless analyzing method according to the situation which could leads to the different level of accuracy of the result according to the data set.

หนึ่งในวิธีการที่ simple และมีการถูกนำมาใช้ใน CRM อย่างแพร่หลายคือการหาความสัมพันธ์ของสินค้าแต่ละชนิดว่ามีโอกาสที่จะถูกซื้อมากเท่าไหร่เมื่อพิจารณาจากสินค้าที่มีอยู่ในตะกร้าปัจจุบัน เรียกวิธีการนี้ว่า market basket analytic โดยใช้ association rule algorithm ด้วยวิธีนี้จะช่วยให้ทางองค์กรสามารถคาดเดาพฤติกรรมการซื้อของลูกค้าเพื่อวางแผนกลยุทธ์ทางการตลาดได้. Market basket analytic นั้นสามารถนำไปประยุคต์ใช้ผ่าน tool และวิธีวิเคราะห์ที่หลากหลาย ในสถานการณ์ที่ต่างกันไป โดยจะให้ผลลัพธ์ที่มีประสิทธิภาพที่ไม่เหมือนกันซะทีเดียวในข้อมูลแต่ละชุด.

This study is focusing on the process of analyzing the association rule which came from the apriori algorithm and the frequent pattern growth algorithm in the different kind of development tool or programming language which includes Weka, Python, Rstudio and Matlab. The result is expected to be the efficiency analyzation and comparison between different process, also the display of the result and the discussion.

การศึกษานี้มุ่งเน้นไปที่การวิเคราะห์ association rule ที่ได้มาจาก apriori algorithm และ frequent pattern growth algorithm ใน development tool หรือ programming language ที่ต่างกันไป ซึ่งประกอบไปด้วย Weka, Python, Rstudio, และ Matlab และนำผลลัพธ์มาวิเคราะห์เพื่อเปรียบเทีบประสิทธิภาพของวิธีการต่างๆ ก่อนที่จะนำไปแสดงผล และอภิปรายผลลัพธ์ในขั้นตอนสุดท้าย

This study is aim to display the efficiency result of the algorithm which includes time consumption and accuracy of the analyzing process of the association rule from various data analytic tool in order to improve the understanding of each tool so that it can be applied in the most sufficient way possible.

การศึกษานี้ตั้งเป้าไปที่การแสดงผลลัพธ์ด้านประสิทธิภาพ ได้แก่ ระยะเวลาที่ใช้ และความแม่นยำ ของการวิเคราะห์ Association rule จาก data analytic tool ประเภทต่างๆ เพื่อให้เกิดความเข้าใจในการเลือกใช้ tools ให้เหมาะสนกับสถานการณ์