

การทบทวนการวิเคราะห์ข้อมูลเชิงทำนายสำหรับข้อมูลขนาดใหญ่ในธุรกิจ

REVIEW OF PREDICTIVE ANALYTICS FOR BIG DATA IN BUSINESS

ชนกานต์ กิ่งแก้ว¹

บทคัดย่อ

การวิเคราะห์ข้อมูลเชิงทำนาย (Predictive Analytics) คือ เทคโนโลยีซึ่งเรียนรู้จากประสบการณ์ (ข้อมูล) เพื่อการทำนายพฤติกรรมบางอย่างที่เกิดขึ้นในอนาคต ซึ่งจะทำให้สามารถช่วยในการตัดสินใจในทางธุรกิจได้ดียิ่งขึ้น ซึ่งประกอบไปด้วยเทคนิคหลายๆ ด้าน ประกอบไปด้วย หลักสถิติ การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล ในทางธุรกิจนั้น การวิเคราะห์ข้อมูลเชิงทำนายเป็นการสร้างโมเดลของรูปแบบซึ่งได้มาจากข้อมูลในอดีตเพื่อหาโอกาส หรือ ความเสี่ยง ซึ่งในแต่ละวันนั้นมีการตัดสินใจเกิดขึ้นอย่างมากมาย ซึ่งการตัดสินใจต่างๆ เหล่านี้ย่อมมีผลกระทบ ในทางธุรกิจไม่มากนักน้อย เทคโนโลยีนี้จะช่วยให้การตัดสินใจของมนุษย์นั้นดีขึ้น ด้วยการชี้แนะจากข้อมูลที่มีอยู่ เพื่อตอบ คำถามที่จะเกิดขึ้นในอนาคตได้ ในทางธุรกิจนั้นการวิเคราะห์ข้อมูลเชิงทำนายใช้ในด้านต่างๆ มากมาย อาทิเช่น ด้านการตลาด การเงิน ประกันภัย โทรคมนาคม ค่าปลีก การแพทย์อุตสาหกรรม และ ในด้านอื่นๆ ในบทความนี้จะ พูดยถึงการนำการวิเคราะห์ข้อมูลเชิงทำนายไปใช้ในทางธุรกิจ ประโยชน์ที่จะเกิดขึ้นกับธุรกิจต่างๆ รวมทั้งได้แสดงกรณี ศึกษาในธุรกิจต่างๆ ที่น่าสนใจเอาไว้อีกด้วย

คำสำคัญ : เหมืองข้อมูล การวิเคราะห์ข้อมูลเชิงทำนาย ข้อมูลขนาดใหญ่ การเรียนรู้ของเครื่อง

Abstract

Predictive Analytics is a technology, related to learning-from-experience (data) for predicting the future behavior of individuals in order to drive better decisions. It is combines a variety of techniques such as statistics, modeling, machine learning and data mining. In business aspect, predictive analytics built pattern models relying historical data or experience to identify opportunities or risks. There are millions of decisions made in each day and these will have affect in business in some way. This technology is the mean to drive personnel decisions, guided by data for answering future questions. Predictive Analytics is used in marketing, financial, insurance, telecommunications, retail, travel, healthcare, industrial and other fields. This article presents how to use predictive analytics on business applications, their benefits and interesting case studies.

Keyword : Data Mining, Predictive Analytics, Big Data, Machine Learning

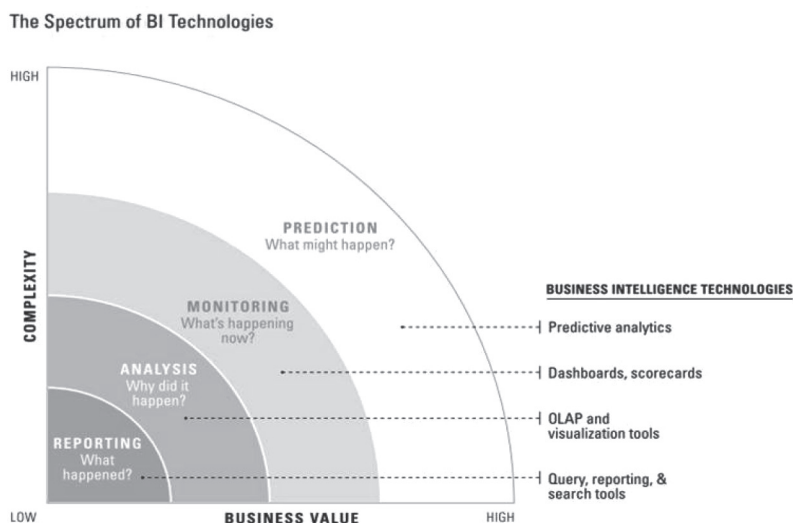
¹ อาจารย์ประจำ สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์
E-mail: chanakarnkin@pim.ac.th

บทนำ

การวิเคราะห์ข้อมูลเชิงทำนายนั้นเป็นแนวคิดที่ได้มาจากการทำเหมืองข้อมูล โดยที่จะจัดการเกี่ยวกับการสกัดข้อมูลที่เป็นประโยชน์ออกมา โดยเฉพาะข้อมูลที่เก็บเอาไว้ในอดีต เพื่อการทำนายแนวโน้มที่จะเกิดขึ้นและรูปแบบของพฤติกรรมต่างๆ โดยในยุคปัจจุบันที่เทคโนโลยีมีการเปลี่ยนแปลงอย่างรวดเร็ว เช่นเดียวกับการแข่งขันในโลกของธุรกิจในปัจจุบันเป็นไปอย่างเข้มข้นจึงเป็นสิ่งที่หลีกเลี่ยงไม่ได้เลยว่าการที่องค์กรจะอยู่รอดได้นั้นจะต้องมีการใช้ข้อมูลที่ทันสมัยและทันท่วงทีเพื่อสนับสนุนการตัดสินใจ (Decision Support) อย่างรวดเร็วและสามารถนำไปวางแผน หรือตอบปัญหาเชิงธุรกิจได้ทันต่อเหตุการณ์ให้กับผู้บริหารระดับสูงขององค์กรในการที่จะได้มาซึ่งข้อมูลเหล่านี้ซึ่งจำเป็นต้องแสวงหาหนทางในการเก็บรวบรวมข้อมูลให้ได้มากที่สุด ทั้งข้อมูลภายในขององค์กรเองและข้อมูลขององค์กรคู่แข่งรวมถึงข้อมูลขององค์กรอื่นๆ ที่อยู่ในการธุรกิจเดียวกัน จากนั้นการเลือกสรรข้อมูลสารสนเทศที่มีคุณค่าจากแหล่งข้อมูลที่มีขนาดใหญ่ (Big Data) เพื่อให้แน่ใจว่าสารสนเทศที่พัฒนาขึ้นมานั้นเป็นสารสนเทศที่สามารถตอบสนองต่อความต้องการของผู้บริหารระดับสูงขององค์กรได้ด้วยเหตุผลดังกล่าวองค์กรจึงจำเป็นต้องมีระบบที่สามารถรวบรวมและวิเคราะห์ข้อมูล เพื่อที่จะได้มาซึ่งสารสนเทศที่มีคุณค่าต่อกิจกรรมทางธุรกิจขององค์กรถ้าการดำเนินการธุรกิจสามารถทำนายปัจจัยต่างๆ เช่น

การดำเนินการ งบประมาณ อุปทานความคิดของลูกค้า ล้วนจะเป็นปัจจัยให้องค์กรประสบความสำเร็จได้เป็นอย่างดี การวิเคราะห์เชิงทำนายนั้นเป็นศาสตร์ที่ช่วยให้ธุรกิจนั้นรู้ว่าลูกค้าจะอย่างไรต่อไป แม้ในปัจจุบันธุรกิจต่างๆ ได้ทำการเก็บข้อมูลที่มีประโยชน์ เช่น การใช้จ่าย การซื้อขาย การทำประกัน การฝากถอนเงิน การสั่งซื้อสินค้า การใช้บัตรเครดิต การใช้เว็บไซต์ การใช้โทรศัพท์ การพบแพทย์ และอื่นๆ อีกมากมาย ได้มีการใช้เครื่องมือการวิเคราะห์ข้อมูลที่รู้จักกันในชื่อของ Business Intelligence หรือ BI นั้นเอง ซึ่งเป็นการนำข้อมูลที่มีอยู่มาจัดทำรายงานในรูปแบบต่างๆ ที่เหมาะสมกับมุมมองในการวิเคราะห์ให้ตรงตามความต้องการขององค์กรเพื่อประโยชน์ในการวางแผนกลยุทธ์ด้านต่างๆ ซึ่งเครื่องมือ BI นั้นสามารถแบ่งออกได้เป็น 4 ระดับ ดังที่แสดงในรูปที่ 1 ซึ่งแสดงถึงระดับของเทคโนโลยี BI ต่อความซับซ้อนของการวิเคราะห์ข้อมูล และคุณค่าที่ได้ต่อธุรกิจ ได้แก่

1. เครื่องมือการสอบถาม การทำการรายงาน และค้นหา (Query, Reporting & Search Tools)
2. กระบวนการวิเคราะห์ผลแบบออนไลน์ และเครื่องมือสร้างมโนทัศน์ (OLAP and Visualization tools)
3. หน้าต่างแสดงข้อมูล (Dashboard and Score-cards)
4. การวิเคราะห์เชิงทำนาย (Predictive Analytics)



รูปที่ 1: ระดับของเทคโนโลยี BI ความซับซ้อนของการวิเคราะห์ข้อมูล ต่อคุณค่าที่ได้ออกมาทางธุรกิจ
ที่มา: (Eckerson, 2007) <http://www.witinc.com/Predictive-Analytics.id.355.htm>

Query, reporting & search tools

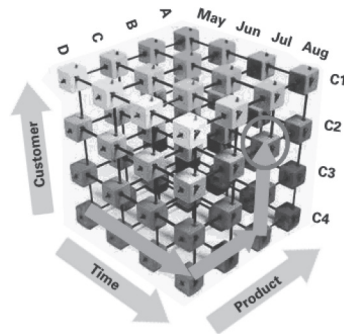
EmployeeID	FirstName	LastName	HireDate	City
5	Steven	Buchanan	17/:0/1993 12:00:00 AM	London
5	Michael	Suyama	17/:0/1993 12:00:00 AM	London
7	Robert	King	2/1/1994 12:00:00 AM	London
9	Anne	Dodsworth	15/:1/1994 12:00:00 AM	London
2	Andrew	Fuler	14/8/1992 :2:00:00 AM	Tacoma
1	Nancy	Devolio	1/5/1992 12:00:00 AM	Seattle
3	Laura	Callahan	5/3/1994 12:00:00 AM	Seattle
4	Margaret	Peacock	3/5/1993 12:00:00 AM	Redmond
3	Janet	Leverling	1/4/1992 12:00:00 AM	Kirkland

รูปที่ 2: แสดงข้อมูลที่ได้จากการสอบถามข้อมูล
ที่มา: technet.microsoft.com

เป็นรายงานที่สร้างขึ้นเฉพาะกิจ เมื่อผู้ต้องการใช้งานร้องขอ เช่น ต้องการจัดลำดับลูกค้าทั้งหมดที่ซื้อผลิตภัณฑ์มากกว่า 5,000 ดอลลาร์ในช่วงเดือนมกราคม 2006 เป็นการสอบถามข้อมูล (Query) โดยใช้ภาษาสืบค้น

ข้อมูลที่อยู่ในรูปแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) โดยใช้ภาษา SQL เพื่อนำข้อมูลออกมาแสดงในรูปแบบของตาราง

OLAP and Visualization tools

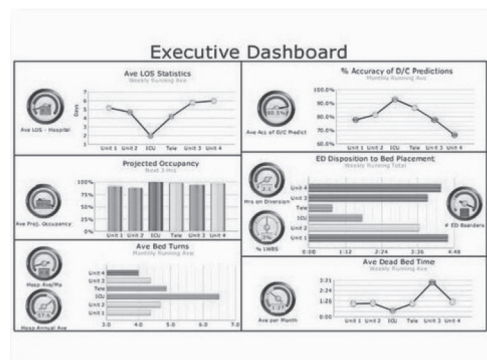


รูปที่ 3: แสดงภาพ OLAP ที่มีการทำข้อมูลออกมาเป็นหลายมุมมอง
ที่มา: <http://www.oracle.com/>

OLAP หรือ Online Analytical Processing เป็นเทคโนโลยีที่ประกอบด้วยเครื่องมือที่ช่วยดึงและนำเสนอข้อมูลในหลายมิติ (Multidimensional) จากหลายๆ

มุมมอง โดยที่ OLAP ได้รับการออกแบบมาสำหรับผู้ใช้ในในระดับของผู้บริหารหรือหน่วยงานในองค์กร ที่ต้องวิเคราะห์ข้อมูลเพื่อใช้ประกอบการตัดสินใจในระดับสูง

Dashboard and Scorecards

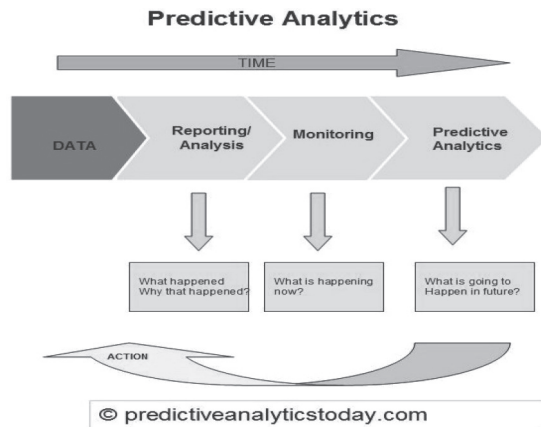


รูปที่ 4: แสดงภาพ Dashboard ที่มีการกำหนดตัวชี้วัด
ที่มา: <http://www.kaushik.net/>

เป็นการแสดงมโนทัศน์ (Visualization) ในด้านที่สำคัญ และจะมีการกำหนดตัวชี้วัด (KPI) ไว้ด้วย เพื่อแสดงถึงประสิทธิภาพขององค์กรจะมีการแบ่งแถบสีแสดงสถานะต่างๆ เพื่อสามารถติดตาม ประมวลผลได้

และตัดสินใจได้ เช่น การวัดยอดขาย มีการกำหนดระดับเพื่อเอาไว้อ้างอิง มีการแบ่งตามมิติภูมิภาค เพื่อช่วยให้ผู้บริหารได้ทำการตัดสินใจได้ดีขึ้นจากข้อมูลที่มีอยู่

Predictive Analytics



รูปที่ 5: ภาพรวมของการวิเคราะห์ข้อมูลเชิงทำนาย
ที่มา: predictiveanalyticstoday.com

การวิเคราะห์ข้อมูลเชิงทำนายเป็นการใช้โมเดลขั้นสูงเพื่อสามารถวิเคราะห์ไปข้างหน้าได้จากข้อมูลในอดีตที่มีขนาดใหญ่ (Big Data) ข้อมูลเหล่านี้จะถูกนำเข้าผ่านกระบวนการต่างๆ การทำให้ข้อมูลแสดงออกมาเป็นรายงานและการวิเคราะห์ (Report/Analysis) จะเป็นการหาคำตอบที่เกิดขึ้นในอดีต เช่น เกิดอะไรขึ้นและทำไมถึงเกิดเช่นนั้น การติดตามผล (Monitoring) เป็นการติดตามผลที่เกิดขึ้นในปัจจุบัน จากนั้นขั้นตอนการใช้การวิเคราะห์ข้อมูลเชิงทำนาย (Predictive Analytics) นั้นจะสามารถตอบคำถามที่ว่า อะไรจะเกิดขึ้นในอนาคต และอะไรที่น่าจะเกิดขึ้นมากที่สุด? การวิเคราะห์จะได้สิ่งที่เกิดขึ้นในอนาคต ซึ่งจะก่อให้เกิดการตัดสินใจ (Action) เพื่อประโยชน์สูงสุดในทางธุรกิจได้ ภาพรวมของการวิเคราะห์ข้อมูลเชิงทำนายแสดงในรูปที่ 5 เริ่มจากการใช้ข้อมูลที่มีอยู่มาทำการวิเคราะห์โดยข้อมูลสามารถแปลงไปเป็นการแสดงรายงาน

จากเทคโนโลยีต่างๆ ที่กล่าวมาในข้อ 1-3 นั้น ล้วนแล้วแต่เป็นการนำเสนอข้อมูลในอดีตเท่านั้นซึ่งจะนำคุณค่ามาให้ในทางธุรกิจได้น้อยกว่าซึ่งเทคโนโลยีการวิเคราะห์เชิงทำนายนั้นจะมาตอบโจทย์ในอนาคตซึ่งจะทำให้มีคุณค่าทางธุรกิจได้สูงกว่า สามารถทำให้บริษัทมีผลตอบแทนจากการลงทุนกว่า 250% (Predictive Analytics World, 2009) เป็นการคาดการณ์ที่ได้จากโมเดลขั้นสูง ซึ่งได้มาจากเทคนิคหลายเทคนิค ได้แก่ การทำเหมืองข้อมูล หลักการทางสถิติ และการเรียนรู้ของจากที่กล่าวมานั้นจะเห็นว่าเทคโนโลยีต่างๆ ได้มีวิวัฒนาการมาจากการเก็บรวบรวมข้อมูลต่างๆ ในรูปแบบดิจิทัลเป็นครั้งแรกในช่วงปี 1960 และ พัฒนาเรื่อยมาจนถึงปัจจุบัน ที่เราได้ทำการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) คือ ข้อมูลที่มีขนาดใหญ่ มีรูปแบบไม่ตายตัวยากต่อการประมวลผล ดังตารางที่ 1

ตารางที่ 1 : วิวัฒนาการของการใช้ข้อมูล

วิวัฒนาการ	คำถามเชิงธุรกิจ	เทคโนโลยีที่ใช้งานได้	ลักษณะเฉพาะ
การรวบรวมข้อมูล (1960)	อะไรคือรายได้รวมห้าปีสุดท้าย	คอมพิวเตอร์ เทป ดิสก์	การส่งข้อมูลคงที่ และส่งผลย้อนหลัง
การเข้าถึงข้อมูล (1980)	อะไรคือหน่วยการขายใน New England เมื่อเดือนมีนาคมที่แล้ว	ฐานข้อมูลเชิงสัมพันธ์ (RDBMS) โครงสร้างภาษาสอบถามข้อมูล (SQL)	ส่งข้อมูลแบบไดนามิก และส่งผลย้อนหลัง
คลังข้อมูลและสนับสนุนการตัดสินใจ (ต้นปี 1990)	อะไรคือยอดขายในพื้นที่ A โดยผลิตภัณฑ์และพนักงานขาย	OLAP ฐานข้อมูลหลายมิติ คลังข้อมูล	ส่งข้อมูลหลายระดับ และผลย้อนหลัง
เหมืองข้อมูลอัจฉริยะ (ปลายปี 1990)	เกิดอะไรขึ้นในหน่วยการขายใน Boston เดือนหน้า และทำไมจึงเป็นเช่นนั้น	วิธีการขั้นสูง Multiprocessor Computer ฐานข้อมูลใหญ่	ส่งข้อมูลที่คาดไว้และข้อมูลเชิงรุก
ระบบอัจฉริยะขั้นสูง การบูรณาการที่สมบูรณ์ (2000-2004)	แผนการที่ตามมาที่ดีที่สุดคืออะไร ทำการเปรียบเทียบการวัดผลได้อย่างไร	การประมวลผลเชิงประสาทโมเดล AI ขั้นสูง การเพิ่มประสิทธิภาพที่ซับซ้อน Web service	คู่ค้าทางธุรกิจที่หลากหลาย

ที่มา: (มหาวิทยาลัยเทคโนโลยีมหานคร, 2009)

ยุคของ Big Data

ข้อมูลขนาดใหญ่ หรือ Big Data สามารถนิยามความหมายของ Big Data ได้ด้วย 3V คือ Volume, Velocity และ Variety กล่าวคือ ขนาดของข้อมูลที่มีขนาดใหญ่มากเกินกว่าฐานข้อมูลแบบเดิมจะสามารถจัดการได้ (Volume) ข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็ว เช่น ข้อมูลจากสังคมออนไลน์ ข้อมูลการซื้อขาย (Velocity) และ โครงสร้างข้อมูลนั้นแตกต่างจากระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) กล่าวคือ ข้อมูลในยุคปัจจุบันนั้นไม่ได้มีโครงสร้างเชิงสัมพันธ์อีกต่อไป มีหลากหลาย เช่น ข้อมูลแบบสตรีม ข้อมูลแบบเรียลไทม์ (Real Time) ข้อมูลจากเครือข่ายสังคมออนไลน์ ข้อมูลแบบข้อความ (Variety) ทำให้เกิดความท้าทายที่เกิดขึ้นจากการใช้ Big Data ได้แก่ การได้มาซึ่งข้อมูล

การเก็บข้อมูลมหาศาล การค้นหาข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล การแสดงมิติทัศน์ของข้อมูล ขนาดใหญ่เป็นเรื่องที่ยากเป็นอย่างยิ่ง โดยเฉพาะในกรณีของเว็บไซต์ต่างๆ โดยเฉพาะเครือข่ายสังคมออนไลน์ที่มีผู้ใช้ในระดับร้อยล้านคนขึ้นไป การซื้อของในร้านสะดวกที่ต้องการการวิเคราะห์ข้อมูลในระดับหมื่นล้านรายการ ข้อมูลข้อความบนเว็บไซต์ต่างๆ และอื่นๆ ทำให้ต้องมีเทคโนโลยีที่มารองรับในการวิเคราะห์ข้อมูลเหล่านี้ โดยในปัจจุบันเทคโนโลยีมีความก้าวหน้าไปอย่างรวดเร็ว การเก็บข้อมูลมีราคาถูกลงเนื่องจากราคาของหน่วยความจำมีความถูกลงอย่างมาก การใช้สถาปัตยกรรมแบบใหม่ การใช้ดีสก์ที่ทำจากวัสดุขั้นดี สามารถทำให้การวิเคราะห์ข้อมูลเป็นไปได้ได้ดี รวดเร็วมมากขึ้น

ขั้นตอนการทำการวิเคราะห์เชิงทำนาย (Predictive Analytics Process)



รูปที่ 6: ขั้นตอนการทำ Predictive Analytics ตามมาตรฐาน CRISP-DM (C. Shearer, 2006)

จากรูปที่ 6 เป็นภาพแสดงมาตรฐานการทำการวิเคราะห์เชิงทำนายตามมาตรฐานแบบ CRISP-DM หรือ Cross Industry Standard Process for Data Mining โดยพัฒนาขึ้นโดยบริษัทชั้นนำ 3 บริษัทอย่าง Daimler AG, SPSS, NCR (Chapman, 2000) ซึ่งประกอบไปด้วย 6 ขั้นตอนมาตรฐาน ดังต่อไปนี้

1. ขั้นตอนการเข้าใจในจุดประสงค์ทางธุรกิจ (Business Understanding) เป็นขั้นตอนที่วิเคราะห์ความต้องการทำเหมืองข้อมูลในเชิงธุรกิจ เมื่อทราบจุดประสงค์แล้วก็ระบุผลลัพธ์ หรือ เป้าหมายที่ต้องการที่จะได้จากการวิเคราะห์ และขั้นตอนนี้ยังเป็นขั้นตอนการวางแผนงานอีกด้วย

2. ขั้นตอนการเข้าใจในข้อมูล (Data Understanding) เป็นขั้นตอนที่มีการรวบรวมข้อมูลที่เกี่ยวข้อง น่าเชื่อถือ ในจำนวนที่มากเพียงพอที่จะนำมาใช้ในการวิเคราะห์ เนื่องจากข้อมูลเป็นปัจจัยที่สำคัญที่สุดในกระบวนการ

3. ขั้นตอนการเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ใช้เวลายาวนานที่สุด เนื่องจากความถูกต้องที่จะได้จากการทำการวิเคราะห์ข้อมูลนั้นขึ้นอยู่กับคุณภาพของข้อมูล ซึ่งประกอบไปด้วยขั้นตอนย่อยๆ อีก คือ การคัดเลือกข้อมูล (Data Selection) เป็นการ

กำหนดเป้าหมายว่าสิ่งที่ต้องการจะวิเคราะห์คืออะไร

- การกลั่นกรองข้อมูล (Data Cleaning) เป็นการทำให้ข้อมูลมีความถูกต้อง โดยจะมีการลบข้อมูลที่ซ้ำซ้อนกัน ซ่อมแซมข้อมูลที่ขาดหายไป รวมไปถึงแก้ไขข้อมูลที่มีข้อผิดพลาด การตัดข้อมูลที่ไม่มีอยู่ในช่วงออกเป็นต้น

- การบูรณาการข้อมูล (Data Integration) เป็นการรวบรวมข้อมูลจากหลายๆ แหล่งเข้าด้วยกัน เนื่องจากในบางครั้งข้อมูลมาจากหลายแหล่ง

- การลดขนาดข้อมูล (Data Reduction) เป็นการลดขนาดข้อมูล ในบางครั้งข้อมูลที่มีมากเกินไปจะทำให้การทำงานในการสร้างแบบจำลองมีความช้าและเสียพื้นที่ในการเก็บข้อมูลอย่างมาก

- การแปลงรูปข้อมูล (Data Transformation) เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมไปใช้ในการวิเคราะห์ข้อมูลตามหลักอัลกอริทึมของการทำเหมืองข้อมูล บางครั้งข้อมูลบางประเภทไม่สามารถทำการสร้างแบบจำลองได้ ยกตัวอย่างเช่น ข้อมูลแบบรายการการซื้อสินค้า ไม่สามารถใช้หลักอัลกอริทึมของกฎความสัมพันธ์ (Association Rule) ได้ ต้องมีการทำการแปลงรูปข้อมูลให้อยู่ในรูปแบบของข้อมูลเชิง

สัมพันธ์ (Relational Data) อีกทั้งข้อมูลประเภท Text ก็ไม่สามารถใช้ในอัลกอริทึมประเภทการจำแนกประเภทของข้อมูลได้ ต้องมีการแปลงรูปข้อมูล ซึ่งในบางครั้งเรียกว่า การสกัดฟีเจอร์ (Feature Extraction) กระทำในขั้นตอนนี้ การแปลงข้อมูลให้อยู่ในช่วงที่กำหนดเอง (Normalization)

- การทำให้ข้อมูลเป็นช่วง (Data Discretization) เป็นการทำให้ข้อมูลที่เป็นตัวเลข อยู่ในจำนวนของช่วงที่ต้องการ

4. ขั้นตอนการสร้างโมเดล (Modeling) เป็นขั้นตอนในการใช้อัลกอริทึมในการสร้างโมเดล

5. ขั้นตอนการตรวจสอบและประเมินผล (Evaluation) เป็นขั้นตอนในการวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลเชิงทำนาย ดูจากความแม่นยำในการทำนาย การสามารถนำโมเดลไปใช้ในทางธุรกิจได้จริง

6. ขั้นตอนการนำไปใช้งาน (Deployment) เป็นขั้นตอนที่นำโมเดลผลที่ได้จากการวิเคราะห์ข้อมูลไปใช้จริง เพื่อช่วยในการตัดสินใจทางธุรกิจ

ประเภทของโมเดล

โดยทั่วไปแล้วเมื่อเราพูดถึงการวิเคราะห์ข้อมูลเชิงทำนายก็มักจะพูดถึงโมเดลเชิงทำนาย (Predictive Model) เพียงอย่างเดียว อย่างไรก็ตามได้มีการแบ่งประเภทของโมเดลออกมาเป็นอีก 3 ประเภทตามการวิเคราะห์ข้อมูลได้ คือ โมเดลเชิงทำนาย (Predictive Model), โมเดลเชิงอธิบาย (Descriptive Model) และโมเดลการตัดสินใจ (Decision Model) โดยโมเดลเชิงทำนายนั้นจะทำหน้าที่ในการหาความสัมพันธ์ในหรือรูปแบบที่ซ่อนตัวอยู่ในข้อมูล ซึ่งจะทำให้ธุรกิจนั้นสามารถหาพฤติกรรมที่จะเกิดขึ้นในอนาคตจากรูปแบบในอดีตได้ ส่วนโมเดลเชิงอธิบายนั้นจะทำการอธิบายลักษณะของข้อมูลและแบ่งออกข้อมูลออกเป็นกลุ่มๆ เพื่ออธิบายลักษณะของข้อมูลนั้นๆ และโมเดลการตัดสินใจนั้นเป็นการใช้เทคนิคในการหาการตัดสินใจเพื่อให้ได้ผลลัพธ์จากการตัดสินใจนั้น เช่น ดูว่าลูกค้าคนไหนที่มีแนวโน้มที่จะตัดสินใจเปลี่ยนการบริการไปสู่คู่แข่ง (Churn Analysis) เป็นต้น ส่วนโมเดลเชิงตัดสินใจเป็นการหาเทคนิคที่ดีที่สุดในการทำนายผลของการตัดสินใจ ตารางที่ 2 สรุปผลการแบ่งประเภทโมเดล

ตารางที่ 2: การแบ่งประเภทโมเดลตามการวิเคราะห์รูปแบบต่างๆ

	Predictive Models	Descriptive Models	Decision Models
ลักษณะการวิเคราะห์	หาความสัมพันธ์ และรูปแบบจากระหว่างตัวแปรต่างๆ	เป็นการหากลุ่มของข้อมูลที่มีลักษณะใกล้เคียงกัน เพื่ออธิบายลักษณะที่เกิดขึ้น	เป็นการหาการตัดสินใจที่ดีที่สุดจากข้อมูลที่เกิดขึ้น ในการตัดสินใจอย่างใดอย่างหนึ่ง
การใช้ตัวแปร	มุ่งเน้นทำนายโดยใช้ตัวแปรเดียว	มุ่งเน้นการใช้ตัวแปรให้เยอะที่สุดเท่าที่จะทำได้	มุ่งเน้นในการหาการตัดสินใจอย่างใดอย่างหนึ่ง
การนำไปใช้	การวิเคราะห์ความต้องการสินค้าของลูกค้า, การวิเคราะห์เครดิต, การวิเคราะห์การฉ้อฉล (fraud), การวิเคราะห์ความล้มเหลวของระบบ	การจัดกลุ่มลูกค้าแบ่งออกตามลักษณะของประชากร เช่น รายได้ เพศ อายุ ความต้องการของผลิตภัณฑ์	การวิเคราะห์สายทางวิกฤติ (critical path), การวางแผนทรัพยากร, การจัดตารางเวลา, วิเคราะห์กฎทางธุรกิจ

ที่มา : (An Oracle White Paper, 2010)

เทคนิคการวิเคราะห์ข้อมูลเชิงทำนาย

วิธีการและเทคนิคการวิเคราะห์เชิงทำนายสามารถแบ่งได้ออกเป็น 3 ประเภท คือ วิธีการแบบการวิเคราะห์การถดถอย (Regression Analysis), วิธีการเรียนรู้ของเครื่อง (Machine Learning), วิธีแบบการหากฎของความสัมพันธ์ (Association Rules Discovery) โดยแสดงรายละเอียดดังหัวข้อ

การวิเคราะห์การถดถอย (Regression Analysis)

เทคนิคการวิเคราะห์การถดถอยเป็นวิธีที่ใช้สมการทางคณิตศาสตร์เพื่อเป็นโมเดลในการในการอธิบายถึงความสัมพันธ์ของตัวแปรที่จะนำมาพิจารณาโมเดลที่ใช้ได้แก่ Linear Regression, Logistic Regression, Multinomial Logistic Regression, Probit Regression, Logit Regression, Time Series, Duration Analysis, Classification and Regression Tree (CART)

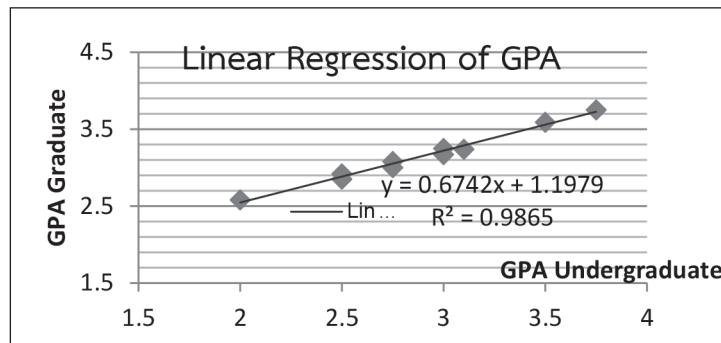
การนำการวิเคราะห์การถดถอยไปใช้ในการวิเคราะห์เชิงทำนาย โดยใช้โมเดล Linear Regression

ตัวอย่างการนำการวิเคราะห์การถดถอยในการคาดการณ์ค่า GPA ของนักศึกษาปริญญาโทที่ได้มาจากค่า GPA ของนักศึกษาคนนั้นในช่วงปริญญาตรี โดยมีข้อมูล 2 ตัวแปร คือ GPA ในช่วงปริญญาตรี เป็นตัวแปรที่รู้ค่า (Predictor) และ GPA ในช่วงปริญญาโทเป็นตัวแปรที่ไม่รู้ค่า ที่ต้องการทำนาย (Response) การใช้ Linear Regression เป็นการสร้างสมการเพื่อแสดงความสัมพันธ์ของตัวแปรอิสระ (Attribute) กับ ตัวแปรตาม (Class) ซึ่งในที่นี้ตัวแปรอิสระ คือ GPA ของช่วงปริญญาตรี และตัวแปรตามหรือคลาส คือ GPA ช่วงปริญญาโท ซึ่งคลาสคำตอบต้องเป็นตัวเลข (Numeric) เท่านั้น ดังตารางที่ 3

ตารางที่ 3 ข้อมูลนำเข้าสำหรับการสร้างโมเดล Linear Regression

เลขที่นักศึกษา	GPA (Undergraduate)	GPA (Graduate)
1	3	3.25
2	2	2.58
3	2.5	2.92
4	3.5	3.59
5	3.75	3.75
6	2.75	3.08
7	2.5	2.85
8	2.75	3.00
9	3	3.17
10	3.1	3.24

ซึ่งสามารถแสดงเป็นรูปแบบกราฟได้ดังรูปที่ 7



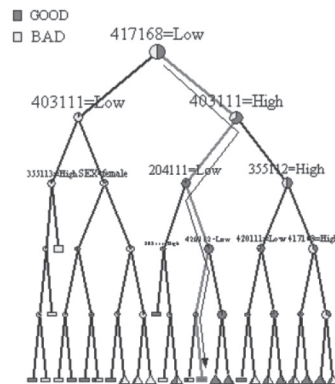
รูปที่ 7: กราฟแสดงโมเดล Linear Regression ระหว่างตัวแปร GPA ปริญญาตรีที่ได้ กับผลของ GPA ปริญญาโท พร้อมสมการเส้นตรงที่เป็นโมเดลอธิบาย

ซึ่งการทำโมเดล Linear Regression จะเป็นการหาสมการเส้นตรง โดยพยายามหาค่าสัมประสิทธิ์ที่ทำให้ Sum Square Error (SSE) มีค่าน้อยที่สุด โดยโมเดลจะอยู่ในรูปแบบสมการ $Y = w_0 + w_1a_1 + w_2a_2 + \dots + w_na_n$ โดยที่ Y คือตัวแปรตาม หรือ คลาสคำตอบ w คือสัมประสิทธิ์ของแต่ละแอตทริบิวต์ a คือ ค่าของข้อมูลในแอตทริบิวต์ที่ 1,2,...n และ n คือ จำนวนของแอตทริบิวต์ทั้งหมด ซึ่งผลลัพธ์ที่ออกมาจะได้สมการ $y = 0.674x + 1.197$ โดยที่ y คือ ค่าประมาณของ GPA ของนักศึกษาปริญญาโท และ x คือ ค่า GPA ของนักศึกษาในระดับปริญญาตรี

การนำไปใช้งานในการทำนาย สมมุติว่ามีนักศึกษาที่เรียนปริญญาตรีได้ GPA คือ 3.0 ดังนั้น เมื่อแทนสมการที่ได้ด้วย 3.0 ก็จะสามารถทำนายได้ว่านักศึกษาคนนี้จะได้ GPA ประมาณ 3.219 นั่นเอง

การนำการวิเคราะห์การถดถอยไปใช้ในการวิเคราะห์เชิงทำนาย โดยใช้โมเดล Classification and Regression Tree

Classification and Regression Tree เป็นโมเดลที่ใช้ในการวิเคราะห์เชิงทำนาย ซึ่งรูปแบบของโมเดลสามารถแปลงออกมาให้เป็นรูปแบบของต้นไม้การตัดสินใจ (Decision Tree) ได้ ตัวอย่างในงานการใช้เทคนิคดาต้าไมนิงเพื่อพัฒนาคุณภาพการศึกษาคณะวิศวกรรมศาสตร์ (กฤษณะ และ คณะ 2003) ได้ทำการสร้างโมเดลต้นไม้การตัดสินใจจากข้อมูลผลการเรียนของนักศึกษาชั้นปีที่ 1 ที่มีผลต่อการเลือกสาขาวิชาของนักศึกษาชั้นปีที่ 2 เพื่อทำนายว่าผลการเรียนของปีที่ 1 มีผลกระทบอย่างไรกับผลการเรียนของปีที่ 2 เพื่อจะได้แนะนำการเลือกสาขาวิชาต่อไป โดยตัวอย่างโมเดลที่ได้จากวิธีนี้จะแสดงออกมาเป็นรูปแบบต้นไม้การตัดสินใจดังรูปที่ 8



รูปที่ 8: โมเดล Classification and Regression Tree แสดงโมเดลต้นไม้การตัดสินใจที่ได้จาก (กฤษณะ และคณะ 2003)

การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่องเป็นวิธีการขั้นสูงในการวิเคราะห์ข้อมูล ซึ่งไม่จำเป็นที่ตัวแปรต้องมีความสัมพันธ์ต่อกัน โดยสามารถนำไปใช้งานได้หลากหลาย เช่น การวินิจฉัยทางการแพทย์, การตรวจสอบการฉ้อโกงบัตรเครดิต, การเรียนรู้จดจำใบหน้าและเสียง และการวิเคราะห์ตลาดหลักทรัพย์เป็นต้น (Wikipedia) โดยโมเดลประเภทในประเภทนี้ได้แก่ Neural Network, Support Vector Machine, Naïve Bayes, K-Nearest Neighbor

การหากฎของความสัมพันธ์ (Association Rules Discovery)

การหากฎของความสัมพันธ์เป็นกระบวนการทำเหมืองข้อมูลที่สำคัญและได้รับความนิยมอย่างมาก บางครั้งจะเรียกว่า Market Basket Analysis หรือการหาความสัมพันธ์ของสินค้าที่ลูกค้าซื้อพร้อมกัน เป็นการหาความสัมพันธ์ของข้อมูล 2 ชุดขึ้นไป ที่มีขนาดใหญ่ โดยมีเทคนิคในการสร้างโมเดลที่นิยม คือ Apriori

ตัวอย่างการหากฎความสัมพันธ์ของ www.amazon.com ดังรูปที่ 9

Special Offers and Product Promotions

- Get \$30 Off: Get the Amazon.com Rewards Visa Card and get an instant \$30.00 Amazon.com Gift Card. Learn more.

Frequently Bought Together





Price for all three: \$205.70

Show availability and shipping details

- ✓ **This item:** Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series ... by Ian H. Witten Paperback \$40.71
- ✓ Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management ... by Jiawei Han Hardcover \$54.71
- ✓ Introduction to Data Mining by Pang-Ning Tan Hardcover \$110.28

Customers Who Bought This Item Also Bought



รูปที่ 9: การใช้กฎความสัมพันธ์ในการตลาด

www.amazon.com ได้มีการสร้างโมเดลกฎของความสัมพันธ์ของสินค้า โดยได้ทำการเก็บข้อมูลการซื้อสินค้าไว้ด้วยกัน จากนั้นแปลงผลเป็นโมเดลในรูปแบบ

กฎแบบ (IF-Then) เช่น ถ้าซื้อหนังสือ Data Mining แล้วลูกค้าก็จะซื้อหนังสือ Predictive Analytics ด้วย การใช้ทางการตลาดของ Amazon คือเมื่อลูกค้าเข้ามา

เลือกซื้อสินค้าใดสินค้าหนึ่งแล้วก็จะแนะนำให้ซื้อสินค้าที่มีการซื้อพร้อมกันด้วย เรียกว่า Product Recommendations โดย Amazon นั้นได้ส่วนแบ่งจากการทำการแนะนำสินค้าแบบนี้กว่าร้อยละ 35 จากยอดขายทั้งหมด (Siegel, 2013)

การนำไปใช้

เรียกได้ว่าบริษัทใหญ่ๆ ในโลกนี้ได้มีการนำการวิเคราะห์เชิงทำนายไปใช้อย่างมากมาย ในชีวิตประจำวัน

เราเองนั้นอาจจะเป็นผู้มีส่วนได้ส่วนเสียโดยตรง โดยในส่วนนี้จะเป็นการแสดงการนำการวิเคราะห์เชิงทำนายไปใช้ในด้านต่างๆ ไม่ว่าจะเป็นด้านครอบครัว ด้านชีวิตส่วนบุคคล การตลาด การโฆษณา การวิเคราะห์ความเสี่ยง การประกันภัย สุขภาพ และด้านอื่นๆ อีกมากมาย ดังตารางที่ 4 ซึ่งแสดงสิ่งที่ได้จากการทำนายและองค์กรที่ได้มีการนำเทคโนโลยีการวิเคราะห์ข้อมูลเชิงทำนายไปใช้นำไปใช้

ตารางที่ 4: การนำการวิเคราะห์เชิงทำนายไปใช้ยังองค์กรต่างๆ (Siegel, 2013)

ทำนายอะไร?	ตัวอย่างการนำไปใช้ขององค์กร
เพื่อน	Facebook: ใช้การแนะนำว่าใครที่น่าจะเป็นบุคคลที่เรารู้จัก (People you may know)
ความรัก	Match.com: ใช้ Intelligent Matching ในการทำนายและจับคู่การเดทระหว่าง 2 คนที่น่าจะสามารถเป็นคนรักกันได้ต่อไป
การตั้งครร์ก	Target: ห้างสรรพสินค้าในอเมริกาสามารถทำนายการตั้งครร์กจากการซื้อสินค้า ทำให้สามารถทำการนำเสนอสินค้าสำหรับแม่และเด็กได้อย่างแม่นยำ
ค่าใช้จ่าย: ในการทำการตลาดเลือกตลาดเป้าหมาย (Target Marketing)	Premier Bankcard: สามารถลดค่าใช้จ่ายในการส่งจดหมายโฆษณาได้กว่า 12 ล้านเหรียญสหรัฐ โดยสามารถส่งได้ตรงเป้าหมายมากขึ้น
การยกเลิก: การทำให้ลูกค้ายังอยู่กับบริการ	Telenor: สามารถลดอัตราการยกเลิกบริการได้กว่า 36%
การขายสินค้า	Amazon.com: 35% ของยอดขายได้มาจากการทำการแนะนำสินค้าจากสินค้าที่ลูกค้าซื้อ
การใช้เว็บไซต์	Google: ปรับปรุงการใช้เครื่องมือค้นหาโดยการทำนายหน้าเว็บที่ผู้ใช้จะเข้าไปใช้งาน Google: ใช้การทำนายว่าอีเมลฉบับไหนจะเป็น spam Netflix: ใช้วิธีการแนะนำภาพยนตร์ออนไลน์ทำให้สามารถดึงดูดการชมภาพยนตร์ออนไลน์ได้มากขึ้นกว่า 70%
การเสียชีวิต	บริษัทประกันภัย: ใช้การทำนายอายุของผู้เอาประกันเพื่อกำหนดเงื่อนไขและราคา
มะเร็งเต้านม	มหาวิทยาลัยสแตนฟอร์ด (Stanford) ได้ทำการใช้โมเดลในการทำนายความเสี่ยงในการเป็นมะเร็งเต้านม ซึ่งได้ผลที่แม่นยำกว่าการวินิจฉัยของแพทย์

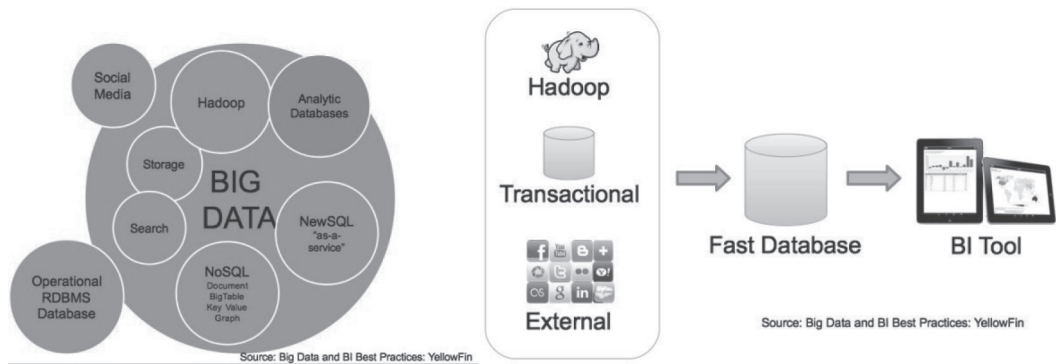
ตารางที่ 4: การนำการวิเคราะห์เชิงทำนายไปใช้ยังองค์กรต่างๆ (Siegel, 2013) (ต่อ)

ทำนายอะไร?	ตัวอย่างการนำไปใช้ขององค์กร
การฉ้อโกง (Fraudulent)	Hewlett-Packard: ประหยัดค่าการรับประกันสินค้าใน 5 ปี กว่า 66 ล้านเหรียญสหรัฐ จากการทำนายการแอบอ้างการส่งคืน สินค้า การเรียกร้องซ่อมแบบผิดปกติของพนักงานขายและลูกค้า Citizen Bank: ทำนายว่าเช็คไหนที่น่าจะเข้าข่ายการโกง โดยสามารถ ลดความเสียหายได้ 20% Aviva Insurance: ปรับปรุงระบบการตรวจจับการขอเรียกร้องจาก ประกันปลอม โดยสามารถประหยัดเงินไปได้กว่า 500,000 ปอนด์ ต่อเดือน
อาชญากรรม	ตำรวจในเมืองชิคาโก ลอสแอนเจลิส เมมphis ริชมอนด์ ซานต้าครูซ ไวน์แลนด์: ได้ใช้การทำนายและสามารถส่งตำรวจไปยังพื้นที่ที่น่าจะ เกิดอาชญากรรมได้ทันเวลา
ความขัดข้องของระบบ	SAS: ใช้ซอฟต์แวร์เพื่อช่วยในการหาการขัดข้องของชิ้นส่วนหนึ่ง ซึ่งจะนำมาซึ่งความขัดข้องของระบบ เช่น ระบบโรงกลั่นน้ำมัน เป็นต้น Argonne National Laboratory: ใช้โมเดลวิเคราะห์เชิงทำนาย ในการหาการขัดข้องของเตาปฏิกรณ์นิวเคลียร์ เพื่อสามารถแก้ไข ปัญหาได้ทัน
การโกหก	University of Buffalo: สามารถสอนให้ระบบเรียนรู้การจับการโกหก จากการเคลื่อนไหวของตา
การใช้พลังงาน	Energex: ทำนายการใช้พลังงานในออสเตรเลียในอีก 20 ปีข้างหน้า Recology: ใช้การวิเคราะห์ข้อมูลเชิงทำนายในการปรับปรุงการ ขนส่งขยะ ซึ่งสามารถลดปริมาณในการส่งขยะไปยังบ่อขยะลงได้ 50%

เครื่องมือที่ใช้ในการพัฒนา

การวิเคราะห์ข้อมูลเชิงทำนายด้วยข้อมูลขนาดใหญ่ (Big Data) จำเป็นจะต้องใช้ระบบการประมวลผลหรือการเก็บข้อมูลในรูปแบบอื่นๆ ที่ไม่ใช่เพียงแบบฐานข้อมูล

เชิงสัมพันธ์ ซึ่งการวิเคราะห์ข้อมูลแบบนี้ต้องใช้โครงสร้างพื้นฐานหลายๆ ด้าน ดังรูปที่ 10 แสดงโครงสร้างพื้นฐานที่สำคัญของการวิเคราะห์ข้อมูล Big Data



รูปที่ 10: โครงสร้างพื้นฐานที่เกี่ยวกับ Big Data
ที่มา: Big Data and BI Best Practices: YellowFinBI

ซอฟต์แวร์ที่สำคัญที่นิยมมาใช้ในการระบบ Big Data คือ Hadoop เป็นซอฟต์แวร์แบบเปิดรหัสที่ได้รับความนิยม จะทำหน้าที่หลักในการแบ่งการประมวลผลและจัดเก็บ ข้อมูลขนาดใหญ่ได้ สามารถเพิ่มขนาดได้อย่างไม่จำกัด โดยจะประกอบไปด้วยส่วนหลัก คือ Hadoop Distributed File System (HDFS) ทำหน้าที่เป็นที่เก็บข้อมูลแบบ ขนานและกระจาย และ ใช้ MapReduce ที่เป็นเทคนิค การประมวลผลข้อมูลแบบขนาน ใช้ครั้งแรกโดย Google เป็นเทคนิคในการเตรียมข้อมูลมาประมวลผลออกเป็น ส่วนย่อยๆ ซึ่งสามารถประมวลผลไปพร้อมๆ กันได้ ผลลัพธ์จากการใช้ Hadoop จะได้ฐานข้อมูลที่มีประโยชน์ ต่อเชิงธุรกิจที่เราสามารถนำซอฟต์แวร์ทางด้านการ วิเคราะห์เชิงทำนายไปใช้ข้อมูลที่ได้จากการประมวลผล ผ่าน Hadoop ต่อไป

เครื่องมือที่ใช้ในการพัฒนาการวิเคราะห์เชิงทำนาย มีแบบทั้งแบบเปิดรหัสและเพื่อการค้า โดยแบบเปิดรหัส ได้แก่ KNIME, Orange, R, Rapid Miner, Weka, GNU Octave, Mahout ส่วนซอฟต์แวร์ทางการค้า ได้แก่ IBM SPSS, MATLAB, SAP, SAS, SAS Enterprise Miner, TIBCO, Oracle Data Mining โดยบริษัท Gartner ซึ่งเป็นบริษัทผู้วิจัยและให้คำปรึกษาด้านเทคโนโลยีชื่อดัง ได้ทำการสำรวจและวิเคราะห์ซอฟต์แวร์ที่ใช้ในการทำ การวิเคราะห์เชิงทำนาย โดยใช้กราฟ Magic Quadrant โดยพิจารณาจากปัจจัยหลักในรูปแบบกราฟ 2 มิติ

แสดงถึงจุดแข็งและความแตกต่างของแต่ละผลิตภัณฑ์ โดยพิจารณาจาก 2 ปัจจัย คือ วิสัยทัศน์ต่อการพัฒนา เทคโนโลยีของผู้ผลิต (Completeness of Vision) ในแกน X และส่วนแบ่งทางการตลาด (Ability to Execute) ในแกน Y ซึ่งจะแบ่งผู้ผลิตซอฟต์แวร์ออกเป็น 4 กลุ่ม ดังรูปที่ 11 แสดง Magic Quadrant ของ ซอฟต์แวร์ที่ใช้ในการวิเคราะห์ข้อมูลเชิงทำนาย

1. Leaders หรือกลุ่มผู้นำในตลาด ซึ่งได้คะแนนสูง ทั้งด้านวิสัยทัศน์ เทคโนโลยี และส่วนแบ่งทางการตลาด ในกลุ่มนี้มี SAS ที่เป็นผู้นำด้านส่วนแบ่งในการตลาด สูงที่สุด และ IBM มีวิสัยทัศน์ด้านเทคโนโลยีมากที่สุด และมี Rapid Miner และ Knime อยู่ในกลุ่มนี้ด้วย

2. Challengers หรือกลุ่มผู้ท้าชิง มีความสามารถในการตอบโจทย์ให้ลูกค้าปัจจุบัน แต่ยังไม่มีความมั่นใจในนวัตกรรมใหม่ๆ ที่เด่นชัด ความเข้าใจในตลาดยังไม่ เพียงพอ

3. Visionaries เป็นกลุ่มที่ตระหนักถึงทิศทางการตลาดและเทคโนโลยีในอนาคต พร้อมที่จะพัฒนา ไปทางนั้นแต่ยังไม่สามารถนำเสนอศักยภาพดังกล่าว ออกมาได้

4. Niche Players หรือกลุ่มตลาดเฉพาะ มีคะแนนต่ำทั้งในด้านวิสัยทัศน์และส่วนแบ่งในตลาด โดยจะมี ลูกค้ากลุ่มเล็กๆ เป็นกลุ่มเฉพาะ เป็นผู้ผลิตเจ้าใหม่ที่กำลัง มาทำตลาด



รูปที่ 11: Magic Quadrant ของซอฟต์แวร์ที่ใช้ในการวิเคราะห์ข้อมูลเชิงทำนาย

ที่มา: <https://www.gartner.com/doc/2667527> - Gartner, 2014

สรุป

ในบทความนี้ได้กล่าวถึงการวิเคราะห์เชิงทำนาย ซึ่งเป็นหลักการที่ใช้ข้อมูลปริมาณมากที่มีอยู่ทั่วไปในองค์กร ซึ่งโดยมากมักมีขนาดใหญ่ มาทำการใช้กระบวนการทางคอมพิวเตอร์ เช่น การทำเหมืองข้อมูล เพื่อให้ได้ออกมาซึ่งโมเดลความรู้ที่สามารถนำไปพัฒนาองค์กรได้ ความรู้ที่ออกมาจากการทำการวิเคราะห์เชิงทำนายนั้นมีค่าเป็นอย่างมาก ซึ่งทำนายสิ่งที่เราต้องการทราบในอนาคตออกมาเพื่อทำการตัดสินใจในทางธุรกิจ ซึ่งจะก่อให้เกิดคุณค่ามากมาย ดังตัวอย่างเช่น บริษัท Amazon ซึ่งเป็นบริษัทขายของออนไลน์ สามารถทำกำไรกว่า 35% เกิดจากการทำการวิเคราะห์สินค้าที่ลูกค้ามีโอกาสซื้อต่อไป บริษัทบัตรเครดิตลดความเสี่ยงในการให้สินเชื่อได้มาก ผู้ให้บริการเครือข่ายโทรศัพท์ สามารถ

เข้าใจลูกค้าและสามารถทำนายการเปลี่ยนบริการของลูกค้าได้ แต่ขั้นตอนในการทำให้ออกมาเป็นผลดังกล่าว เป็นสิ่งที่ยากยิ่งกว่า เริ่มจากการกำหนดปัญหาที่ต้องการทราบทางธุรกิจ การทำความเข้าใจกับข้อมูลที่มีอยู่ การเตรียมข้อมูล การสร้างโมเดล การนำมาใช้งาน การประเมินผลการใช้งานของโมเดล ซึ่งมีเครื่องมือให้เลือกใช้อย่างมากมาย ในตอนนี้หลายๆ บริษัทชั้นนำของโลกได้นำวิธีการวิเคราะห์เชิงทำนายไปใช้อย่างแพร่หลาย สามารถทำให้บริการตอบโต้ลูกค้ามากขึ้น นำเสนอสินค้าได้ตรงความต้องการ ลดค่าใช้จ่ายในการดำเนินการได้อีกด้วย ทำให้เป็นเทคโนโลยีที่น่าจับตามองมากในปัจจุบัน องค์กรที่นำวิธีการวิเคราะห์ข้อมูลเชิงทำนายก็จะสามารถดำเนินธุรกิจไปได้ด้วยดี

บรรณานุกรม

- กฤษณะ ไวยมัย, ชิตชนก ส่งศิริ และ ธนาวิทย์ รักธรรมานนท์. (2003). การใช้เทคนิคดาต้าไมนิงเพื่อพัฒนาคุณภาพ การศึกษาคณะวิศวกรรมศาสตร์. *NECTEC Technical Journal Vol. III, No. 11* 134.
- Collin Shearer. (2006). *First CRISP-DM 2.0 Workshop Help*. KDNuggets สืบค้นเมื่อ 25 มีนาคม 2557, จาก <http://www.kdnuggets.com/news/2006/n19/4i.html>
- Eric Siegel. (2005). *Predictive Analytics with Data Mining: How It works*. *DM Review's DM Direct*. สืบค้นเมื่อ 10 มีนาคม 2557, จาก <http://www.predictionimpact.com/predictive.analytics.html>
- Eric Siegel. (2013). *Predictive Analytics: The power to predict who will click, buy, lie or die book (1st ed.)*. New Jersey: Wiley.
- Gartner. (2014). Magic Quadrant for Advanced Analytics Platforms (2014). Retrieved March 10, 2014, from <https://www.gartner.com/doc/2667527>
- Oracle. (2010). Predictive Analytics: Bringing The Tools To The Data. An Oracle White Paper.
- Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth. (2000). *CRISP-DM 1.0 Step-by-step data mining guides*, สืบค้นเมื่อ 20 กุมภาพันธ์ 2557, จาก <ftp://ftp.software.ibm.com/CRISP-DM.pdf>
- Wayne W. E. (2007). *Predictive Analytics - Extending the Value of Your Data Warehousing Investment*. First Quarter 2007 - TDWI Best Practices Report. Retrieved March 23, 2014, from www.sas.com/events/cm/174390/assets/102892_0107.pdf



Chanakarn Kingkaew received his Bachelor Degree of Engineering in Software and Knowledge Engineering from Faculty of Engineering, Kasetsart University in 2008 with outstanding educational record. In 2010, he gained the professional experience as Research Student in Ericsson GmbH Eurolab in Aachen, Germany and also an Exchange Student in Germany for 2 years. In 2011, he graduated Master Degree of Science in Software Systems Engineering from Thai-German Graduate School of Engineering, King Mongkut University of Technology, North Bangkok in Thailand. At present, he is a full-time lecturer in Faculty of Engineering and Technology, Panyapiwat Institute of Management and IT advisor of CP Retailink.