## Text and Web Mining: Text and Web Mining, Social Analytics, Sentiment Analysis

1. List and briefly discuss some of the text mining applications in marketing.

Text mining can be used to increase cross-selling and up-selling by analyzing the unstructured data generated by call centers.

Text mining has become invaluable for customer relationship management. Companies can use text mining to analyze rich sets of unstructured text data, combined with the relevant structured data extracted from organizational databases, to predict customer perceptions and subsequent purchasing behavior.

1. What are the main steps in the text mining process?

See Figure 7.5 (p. 304). Text mining entails three tasks:

* Establish the Corpus: Collect and organize the domain-specific unstructured data
* Create the Term–Document Matrix: Introduce structure to the corpus
* Extract Knowledge: Discover novel patterns from the T-D matrix

1. What is Web content mining? How does it differ from text mining?

Web content mining refers to the extraction of useful information from Web pages. The documents may be extracted in some machine-readable format so that automated techniques can generate some information about the Web pages.

1. Define Web structure mining, and differentiate it from Web content mining.

Web structure mining is the process of extracting useful information from the links embedded in Web documents.

1. Define Web usage mining.

* Web usage mining is the extraction of useful information from data generated through Web page visits and transactions. Including Clickstream analysis, Log data in server access logs, referrer logs, agent logs and client-side cookies

1. What is meant by social analytics

Social analytics involves monitoring, analyzing, measuring, and interpreting digital interactions and relationships of people, topics, ideas, and content. It involves mining the textual content created in social media (e.g., sentiment analysis, natural language processing) and analyzing socially established networks (e.g., influencer identification, profiling, prediction). A social network is a social structure composed of individuals/people (or groups of individuals or organizations) linked to one another with some type of connections/relationships. Social network analysis (SNA) is the systematic examination of social networks, and is an interdisciplinary field that emerged from social psychology, sociology, statistics, and graph (network) theory. Social analytics therefore combines text analysis for content and sentiment in online communications with social network analysis to identify and analyze relationships between individuals in a community.

Social media analytics refers to the systematic and scientific ways to consume the vast amount of content created by Web-based social media outlets, tools, and techniques for the betterment of an organization’s competitiveness. It is done using many analytic methods, including text mining, sentiment analysis, and social network analysis. Companies use it to get a better understanding of their customer base, and can gain financial and competitive advantages from doing so. Governments use it to track potential terrorist threats, which can lead to enhanced national security. Social scientists use it to get a better understanding of how communities and societies work, which can provide guidance on how to best manage these societies.

1. What is a social network?

A social network is a social structure composed of individuals/people (or groups of individuals or organizations) linked to one another with some type of connections/relationships. Social network analysis (SNA) is the systematic examination of social networks, and is an interdisciplinary field that emerged from social psychology, sociology, statistics, and graph (network) theory. Social analytics therefore combines text analysis for content and sentiment in online communications with social network analysis to identify and analyze relationships between individuals in a community.

1. How can sentiment analysis be used in predicting financial markets?

How can sentiment analysis be used in predicting financial markets? Many financial analysts believe that the stock market is mostly sentiment driven, so use of sentiment analysis has much relevance for financial markets. Automated analysis of market sentiments using social media, news, blogs, and discussion groups can help with predicting market movements. If done correctly, sentiment analysis can identify short-term stock movements based on the buzz in the market, potentially impacting liquidity and trading.

1. How might we use (a) Supervised and (b) unsupervised learning to detect credit card fraud? (Answer from A Descriptive Study of Credit Card Fraud PatternGlobal Business Review September 201314: 373-384)

Supervised Methods In supervised statistical methods, estimated statistical models are used to discriminate between fraudulent and non-fraudulent purchase behaviour to classify new observations into appropriate classes, say fraudulent and non-fraudulent transactions (Bolton and Hand, 2001). The performance of models can be assessed by measuring their accuracy in correctly classifying new observations as fraudulent or nonfraudulent transactions. Unsupervised methods attempt to detect unusual observations such as customers, transactions, or accounts whose behaviour may be different from the norm. These unusual observations, different from the baseline normal behaviour, are identified for closer examination and subsequent classification. Unlike supervised methods, unsupervised methods do not require samples of fraudulent and legitimate transactions. Hence, unsupervised methods may find use in situations where there is no prior knowledge of fraudulent and legitimate classes of transactions. The other advantage of unsupervised method over supervised method is that previously undiscovered frauds can also be detected

<http://www.forbes.com/sites/kumesharoomoogan/2015/08/06/how-quant-traders-use-sentiment-to-get-an-edge-on-the-market/#3216160c2fd8>

1. What is clickstream analysis? What is it used for?

Analysis of the information collected by Web servers can help us better understand user behavior. Analysis of this data is often called clickstream analysis. By using the data and text mining techniques, a company might be able to discern interesting patterns from the clickstreams.