**Content**

[1. CRM 2](#_Toc20146105)

[1. Role of data mining in CRM 2](#_Toc20146106)

[2. Data integrity 5](#_Toc20146107)

[3. Ad-hoc analysis 5](#_Toc20146108)

[4. Weka 3 5](#_Toc20146109)

[5. SQL v Apache Hive v PIG 5](#_Toc20146110)

[6. Visualization: ggplot2 v D3.js 5](#_Toc20146111)

1. CRM

Customer relationship management (CRM) is an approach to manage a company's interaction with current and potential customers. It uses data analysis about customers' history with a company to improve business relationships with customers, specifically focusing on customer retention and ultimately driving sales growth.

One important aspect of the CRM approach is the systems of CRM that compile data from a range of different communication channels, including a company's website, telephone, email, live chat, marketing materials and more recently, social media. Through the CRM approach and the systems used to facilitate it, businesses learn more about their target audiences and how to best cater to their needs.

* 1. Role of data mining in CRM

**Role of Data Mining in CRM:**

Although it is still a relatively new technology, businesses from all industry verticals be it healthcare, manufacturing, financial, transportation etc. have invested in it to take advantage of historical data. Data mining techniques in CRM can assist your business in finding and selecting the relevant information that may then be used to get a holistic view of the customer life-cycle which comprises of four stages: customer identification, customer attraction, customer retention and customer development. The more data there is in the database, the better the models will be created whose use will result in more business value.

Data mining typically involves the use of predictive modeling, forecasting and descriptive modelling techniques as its key elements. Using these techniques, an organization is able to manage customer retention, select the right prospects & customer segments, set optimal pricing policies, and objectively measure and rank which suppliers are best suited for their needs.

**Applications of Data Mining in CRM**

**1. Basket Analysis**

Ascertain which items customers tend to purchase together. This knowledge can improve stocking, store layout strategies and promotions.

**2. Sales forecasting**

Examining time-based patterns helps businesses make stocking decisions. Furthermore, it helps you in supply chain management, planning your financials and gives complete control over internal operations.

**3. Database marketing**

Retailers can design profiles of customers based on their demographics, tastes, preferences, buying behavior etc. It will also aid the marketing team in designing personalized marketing campaigns and promotion offers.

This will result in enhanced productivity, optimal allocation of company’s resources and bring the desired ROI.

**4. Predictive life-cycle management**

Data mining helps an organization predict each customer’s lifetime value and to service each segment appropriately.

**5. Customer segmentation**

Learn which customers are interested in purchasing your products and design your marketing campaigns and promotions keeping their tastes and preferences in mind. This will increase the efficiency and result in the desired ROI since you won’t be targeting customers who show little to no interest in your product.

**6. Product Customization**

Manufacturers can customize products according to the exact needs of customers. In order to do this, they must be able to predict which features should be bundled to meet the customer demand.

**7. Fraud detection**

By analyzing past transactions that were later determined to be fraudulent, a business can take corrective measures and stop such events from occurring in the future. Banks and other financial institutions will benefit from this feature immensely.

**8. Warranties**

Manufacturers need to predict the number of customers who will submit warranty claims and the average cost of those claims. This will ensure efficient and effective management of company funds.

**Data Mining techniques in CRM**

**Anomaly Detection**

Searching for information that doesn’t match expected behavior or a projected pattern is called anomaly detection. Anomalies can provide actionable information because they deviate from the average in the dataset.

**Association Rule Learning**

Discover relations between data items in huge databases. With association rule learning, hidden patterns can be uncovered and the information gained may be used to better understand customers, learn their habits, and predict their decisions.

**Clustering**

Identify similar data sets and understand both the similarities and differences within the data. Data sets that have similar traits can be used for conversion rate increases. For example, if the buying behavior of one group of customers is similar to that of another group, they can both be targeted with similar services or products.

**Classification**

This technique is used for gathering information about data so that the data sets can be placed into proper categories. One example is the classification of email as either regular, acceptable email or as spam.

**Regression**

Regression analysis is one of the advanced data mining techniques in CRM. The objective is to find the dependency between different data items and map out which variables are affected by other variables. This technique is used to determine customer satisfaction levels and their impact on customer loyalty.

Conclusion:

Data mining is a method or tool that can aid companies in their quest to become more customer-oriented. The combination of CRM and Data Mining tools will augment the knowledge and understanding of customers, products and transactional data, thereby improving strategic decision making and tactical marketing activity. The outcome? Increased revenue due to improved ability to respond to each individual contact and reduced costs due to optimal allocation of resources.

* 1. Data integrity

Data integrity is the maintenance of, and the assurance of the accuracy and consistency of, data over its entire life-cycle, and is a critical aspect to the design, implementation and usage of any system which stores, processes, or retrieves data. The term is broad in scope and may have widely different meanings depending on the specific context – even under the same general umbrella of computing. It is at times used as a proxy term for data quality, while data validation is a pre-requisite for data integrity. Data integrity is the opposite of data corruption. The overall intent of any data integrity technique is the same: ensure data is recorded exactly as intended (such as a database correctly rejecting mutually exclusive possibilities,) and upon later retrieval, ensure the data is the same as it was when it was originally recorded. In short, data integrity aims to prevent unintentional changes to information. Data integrity is not to be confused with data security, the discipline of protecting data from unauthorized parties.

* 1. Ad-hoc analysis

<https://www.sisense.com/glossary/ad-hoc-analysis/>

* 1. Weka 3

Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

* 1. SQL v Apache Hive v PIG

Apache Hive is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis. Hive gives a SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the MapReduce Java API to execute SQL applications and queries over distributed data.

<https://www.whizlabs.com/blog/hive-vs-pig-vs-sql/>

* 1. Visualization: ggplot2 v D3.js