Web Mining Applications in the Education Industry: A CRM Perspective

Ms. Swapna Tekale
Asst. Prof., School of Information Technology, Indira University, Pune

DOI: 10.29322/IJSRP.16.02.2026.p17018
https://dx.doi.org/10.29322/IJSRP.16.02.2026.p17018

Paper Received Date: 10th January 2026
Paper Acceptance Date: 5th February 2026
Paper Publication Date: 12th February 2026

Abstract

The World Wide Web has evolved into an enormous and continuously expanding digital ecosystem characterized by complexity and diversity. Its rapid growth has created significant opportunities for applying analytical approaches derived from data mining, collectively known as web mining. Web mining adapts conventional data mining techniques—such as clustering, classification, association rule discovery, and sequential pattern analysis—to data generated through web environments. Depending on the nature of the data under consideration, web mining is generally categorized into web content mining, web structure mining, and web usage mining.

This paper provides an introductory overview of web mining, with particular emphasis on web usage mining. It further examines the relevance of web mining techniques in the education sector from a Customer Relationship Management (CRM) perspective. The study highlights how educational institutions and EdTech organizations can utilize web-based data to attract, engage, and retain students. By analyzing online behavioral data, institutions can support personalized learning initiatives and make more informed strategic and operational decisions.

Keywords: Web Mining, Data Mining, World Wide Web, Web Usage Mining, Educational CRM

1. Introduction

The World Wide Web represents the largest and most diverse information repository in the digital age. Its dynamic, heterogeneous, and continuously expanding nature makes the extraction of relevant and meaningful information a challenging task. As a response to this challenge, web mining has emerged as a specialized field that applies data mining techniques to web-based data in order to uncover patterns, relationships, and actionable knowledge.

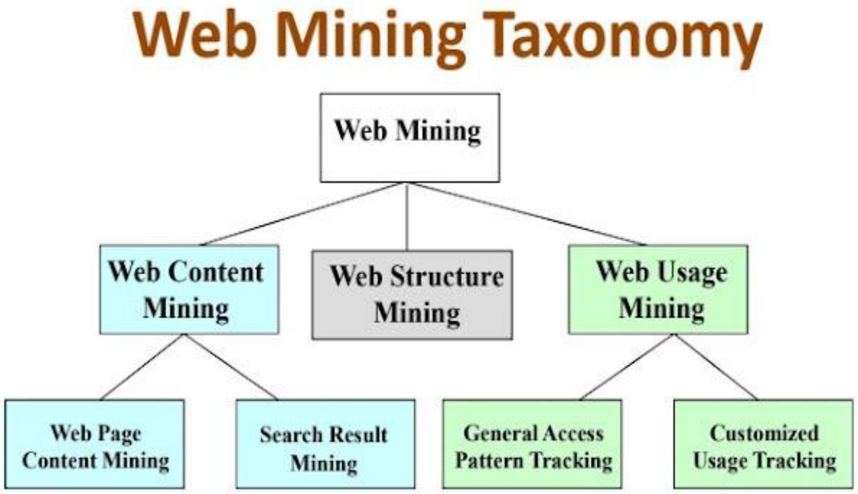
Web mining combines traditional data mining methodologies with data collected from web environments. The concept of "mining" refers to extracting valuable insights from vast amounts of raw, unstructured, or semi-structured data, similar to the extraction of valuable resources from natural deposits. Given the massive volume and velocity of data produced on the web, automated and intelligent analytical tools are essential for effective information discovery.

Organizations apply web mining to gain insights into user behavior, assess website performance, evaluate the effectiveness of marketing initiatives, and support strategic decision-making. Techniques such as clustering, classification, association rule mining, and sequential pattern analysis form the foundation of many web mining applications.

2. Web Mining Taxonomy

Web mining can be broadly categorized into three main types based on the nature of web data being analyzed:

1. Web Content Mining
2. Web Structure Mining
3. Web Usage Mining



2.1 Web Content Mining

Web content mining involves extracting useful information from the actual contents of web documents. Content data refers to the information presented to users on web pages and may include text, images, audio, video, and structured elements such as tables and lists.

This form of mining is often carried out using data collected by search engines and web crawlers. Analytical techniques from artificial intelligence—such as information retrieval, natural language processing, text mining, image analysis, and computer vision—are commonly employed to process and interpret web content.

2.2 Web Structure Mining

Web structure mining focuses on analyzing the structural relationships among web pages by examining hyperlinks. The web can be modeled as a graph in which individual web pages act as nodes and hyperlinks represent edges connecting them.

Web structure mining primarily examines two components:

a) Hyperlink Analysis

Hyperlinks connect one web page to another or link different sections within the same page.

Intra-document links connect different sections of a single page.

Inter-document links connect multiple web pages.

b) Document Structure Analysis

The internal organization of a web page is defined by HTML or XML tags, which form a hierarchical structure known as the Document Object Model (DOM). Analyzing this structure helps in understanding content organization and supports effective information extraction.

2.3 Web Usage Mining

Web usage mining focuses on discovering patterns of user behavior by analyzing interaction data recorded in web logs. It examines how users navigate websites, which pages they visit, and how frequently they interact with specific resources.

Web servers automatically capture data such as IP addresses, timestamps, requested pages, and session details. Analyzing this information helps organizations understand user preferences, improve website design, and enhance overall user experience.

The web usage mining process typically involves three stages:

1. Data preprocessing
2. Pattern discovery
3. Pattern analysis

Types of Usage Data

- **Web Server Data:** Logs generated at the server level
- **Application Server Data:** Business events captured by application servers
- **Application-Level Data:** Custom events recorded by specific applications

Usage data can be collected from:

- Server-side sources
- Client-side sources
- Proxy servers

3. Customer Behaviour Analysis Using Web Mining

Customer Relationship Management (CRM) is one of the most prominent applications of web mining. Websites must be designed not only to attract users but also to retain them and convert them into loyal customers. Web mining techniques analyze visitor behaviour and help predict future interactions, enabling organizations to improve website performance and recommend relevant products or services.

Users exhibit varied behaviours when visiting a website. Some may browse casually, while others may complete transactions such as online banking or purchases. Understanding these behavioural patterns is essential for optimizing website functionality. Web metrics provide standardized measures to evaluate user engagement and site performance.

Term	Description
User	An individual visiting the website
Repeat Visitor	A user who has visited the site more than once
Visitor Recency	Time elapsed between successive visits
Committed Visitor	A visitor spending more than 15 minutes on the site
Stickiness	Frequency and duration of visits
Slipperiness	Ability to quickly find information and exit
Focus	Number of pages visited within a specific section
Speed	Transition rate from visitor to customer

Increasing the number of repeat and committed visitors is a key objective. Customer data, including personal details provided during login, is stored in databases and can be used to personalize content delivery. By analyzing user behavior, organizations can recommend new products, perform market basket analysis, and enhance cross-selling strategies.

Visitor behavior data is commonly obtained from web logs, including:

Log Type	Purpose
Access Logs	Record page requests and hits
Referrer Logs	Track navigation paths
Agent Logs	Identify browser and device details

Analyzing navigation paths before purchases helps identify behavioral patterns and user clusters. Feedback mechanisms further enhance data collection, enabling organizations to refine customer engagement strategies and convert visitors into customers.

4. Web Mining In Education Industry

Educational institutions and EdTech platforms generate significant volumes of web-based data through digital learning environments. Applying web mining techniques to this data supports Student Relationship Management (SRM) and improves academic and administrative services.

4.1 Data Sources in the Education Industry

Common sources of educational web data include:

- Online admission systems
- Learning Management Systems (LMS)
- Student information portals
- Online assessments and assignments
- Course registration platforms
- E-learning systems
- Website navigation logs
- Mobile learning applications

These data sources form the basis for applying web mining techniques in educational contexts.

5. Application of Web Mining Techniques in Education

5.1 Clustering and Classification

Students can be grouped based on academic performance, learning interests, browsing behavior, demographic characteristics, and qualifications. Classification methods help identify students suitable for advanced programs, certifications, scholarships, or education loans.

Example: Students frequently accessing MBA or Data Science course pages can be targeted for postgraduate programs.

5.2 Association Rule Mining

Association rule mining identifies relationships among courses and learning resources, supporting the design of bundled academic offerings.
Example: Students enrolling in Python courses often pursue Data Science programs, enabling institutions to offer combined packages.

5.3 Path Analysis

Path analysis examines student navigation patterns on educational websites, helping identify frequently visited pages and points where users disengage.
Example: If users frequently exit after viewing fee details, institutions can redesign the page or offer flexible payment options.

5.4 Sequential Pattern Mining

This technique analyzes sequences of student activities over time to predict future educational requirements.

Example: A student progressing from a diploma to certification and then to a degree program can be proactively offered advanced courses.

5.5 Data Cube (OLAP) Analysis

Data cube analysis supports multidimensional evaluation using dimensions such as course, academic year, region, and student category, along with measures like enrollment numbers and completion rates.

5.6 AI-Based Techniques

Artificial intelligence approaches, including multi-agent systems and swarm intelligence, support intelligent recommendation engines, adaptive learning environments, and chatbot-based counseling services.

5.7 Web Personalization in Education

Web personalization customizes educational content based on individual student behavior. The process includes:

1. Collecting student-related data
2. Analyzing learning preferences and behavior
3. Delivering personalized recommendations and services

6. Web Mining Applications In Educational CRM

Web mining enables various CRM-related functions in education, including:

- Predicting student dropout rates
- Targeted promotion of academic programs
- Scholarship and education loan recommendations
- Website usability improvement
- Student lifetime value analysis
- Fraud detection in admissions and assessments
- Personalized learning experiences

6.1. Proposed Method

The proposed method presents a structured framework for integrating web mining techniques with an educational Customer Relationship Management (CRM) system to enhance student engagement, inquiry management, and enrollment decision-making. The method transforms raw web interaction data into actionable knowledge that supports personalized and automated CRM workflows.

6.2 Data Acquisition Layer

User interaction data is collected from the institutional website, including page visits, navigation paths, session duration, inquiry submissions, and content downloads. This data is automatically captured through web server log files and online forms and stored in a centralized repository for further processing.

6.3 Data Preprocessing and Session Identification

The collected web data undergoes preprocessing to improve quality and reliability. Noise such as bot traffic, duplicate records, and irrelevant requests is removed. Unique users and sessions are identified using IP address, timestamp, and browser information. The cleaned data is then structured into meaningful user sessions suitable for mining operations.

6.4 Web Mining Module

The core of the proposed method is the web mining module, which applies web usage mining techniques to discover hidden behavioral patterns. Clustering algorithms are employed to segment users based on browsing behavior, interest level, and interaction intensity. Association rule mining is used to identify frequent navigation patterns and relationships between visited pages and inquiry submissions.

6.5 Knowledge Extraction and Pattern Interpretation

The discovered patterns are analyzed to extract knowledge related to user intent, program preferences, and engagement level. These insights are transformed into user profiles representing prospective students, active applicants, and casual visitors. The extracted knowledge serves as input for decision-making within the CRM system.

6.6 CRM Integration and Automation

The interpreted knowledge is integrated into the educational CRM system to enable automated and personalized actions. Based on user profiles and behavioral patterns, the CRM system triggers targeted email campaigns, personalized course recommendations, follow-up notifications, and counselor alerts. This automation ensures timely and relevant interaction with prospective students.

6.7 Performance Evaluation and Feedback Loop

The effectiveness of the proposed method is evaluated using key performance indicators such as inquiry rate, conversion rate, engagement level, and response rate. The evaluation results are fed back into the system to refine mining rules, update user profiles, and improve CRM strategies, thereby creating a continuous learning loop.

6.8 Summary of the Proposed Method

The proposed method establishes an end-to-end framework that integrates web mining with educational CRM workflows. By systematically collecting, mining, and utilizing web interaction data, the approach supports data-driven student relationship management and improves institutional performance.

6.9 Experimental Workflow (Textual Flow Diagram)

Web Server Logs
↓
Data Preprocessing
(Session Identification, Cleaning)
↓
Web Usage Mining
(Clustering & Association Rules)
↓
Pattern Analysis
(User Interest Identification)
↓
CRM Integration
(Personalized Communication)
↓
Performance Evaluation
(Conversion & Engagement Metrics)

6.10 Web Mining Algorithm Used

Algorithm 1: Web Usage Mining for Educational CRM

Input: Web server log files, user interaction data

Output: Student behavior patterns and CRM action triggers

Steps:

1. Collect raw web log data from the institutional website
2. Remove irrelevant records (bots, errors, duplicate entries)
3. Identify unique users and sessions
4. Extract features such as visited pages, time spent, and navigation paths
5. Apply clustering to group users based on browsing behavior
6. Generate association rules to identify frequent navigation patterns
7. Map identified patterns to CRM actions (emails, alerts, recommendations)
8. Monitor user response and update CRM records

6.11. Case Study: Ednet

Experimental Environment

The experiment was conducted using data collected from the official website of a private higher education institution over three consecutive admission cycles.

Data Sources:

- Web server log files
- Online inquiry and application forms
- CRM interaction records

Tools and Techniques Used:

- Web Usage Mining
- Data preprocessing (session identification, noise removal)
- Clustering and Association Rule Mining
- CRM automation modules

Table 1: Dataset Description

Parameter	Description
Observation Period	18 months
Number of Web Sessions	52,000+
Unique Users	31,500
Data Type	Clickstream, page visits, session duration
CRM Actions	Email campaigns, follow-ups, recommendations

Table 2: Performance Comparison (Before and After Implementation)

6.13 Experimental Results

After deploying the web mining–enabled CRM system, institutional performance was measured using key indicators

Metric	Before Implementation	After Implementation
Average Website Bounce Rate	58%	41%
Inquiry Submission Rate	12%	34%
Inquiry-to-Enrollment Conversion	28%	43%
Email Response Rate	18%	37%

6.14 Result Analysis

The experimental results indicate a significant improvement in student engagement and enrollment outcomes after integrating web mining with the CRM system. User clustering enabled targeted communication, while association rule mining helped identify high-intent prospective students. Personalized CRM actions resulted in higher response rates and improved conversion efficiency.

7. Conclusion

Web mining has become an essential tool for extracting meaningful insights from web-based data. Similar to how financial institutions use web mining to target customers, educational institutions can apply these techniques to attract and retain students for courses, certifications, scholarships, and education loans. This data-driven approach enhances student engagement, increases enrollment and retention, supports personalized learning, and strengthens institutional decision-making. Ultimately, web mining enables a shift from mass education to customized education, positioning it as a critical technology for the future of digital learning and educational CRM.

REFERENCES

1. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.
2. J. Srivastava, R. Cooley, M. Deshpande, and P. N. Tan, "Web usage mining: Discovery and applications," ACM SIGKDD Explorations Newsletter, vol. 1, no. 2, pp. 12–23, 2000.
3. R. Kosala and H. Blockeel, "Web mining research: A survey," ACM SIGKDD Explorations Newsletter, vol. 2, no. 1, pp. 1–15, 2000.
4. M. Spiliopoulou, "Web usage mining for web site evaluation," Communications of the ACM, vol. 43, no. 8, pp. 127–134, 2000.

5. B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2nd ed. Berlin, Germany: Springer, 2011.
6. A. Mobasher, R. Cooley, and J. Srivastava, “Automatic personalization based on web usage mining,” Communications of the ACM, vol. 43, no. 8, pp. 142–151, 2000.
7. P. Brusilovsky and M. T. Maybury, “From adaptive hypermedia to the adaptive web,” Communications of the ACM, vol. 45, no. 5, pp. 30–33, 2002.
8. S. Romero and S. Ventura, “Educational data mining: A survey from 1995 to 2005,” Expert Systems with Applications, vol. 33, no. 1, pp. 135–146, 2007.
9. R. Agrawal and R. Srikant, “Mining sequential patterns,” in Proc. 11th Int. Conf. Data Engineering, Taipei, Taiwan, 1995, pp. 3–14.
10. T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.