**1. INTRODUCTION**

**1.1 Problem Definition**

E-commerce businesses face significant challenges in understanding and engaging their customers due to the anonymity of web visitors. Unlike traditional brick-and-mortar stores, where businesses can easily profile their target audience demographically and psychographically, online merchants struggle to gain insights beyond basic sales data.

As web interactions are anonymous, businesses cannot effectively trace or understand visitor behaviors, making it harder to personalize experiences, retain customers, or optimize marketing strategies. Moreover, vast amounts of data are generated daily by e-commerce platforms, which, although valuable, often remain underutilized without appropriate tools for analysis.

Thus, a solution is needed to track online visitor behaviors and analyze web traffic to extract actionable insights that can help businesses improve customer engagement, increase sales, and develop personalized marketing approaches. This requires the development of web traffic tracking and analysis tools that not only collect basic user information (e.g., browsers used, pages visited) but also support advanced analyses such as user path tracking, clustering, and conversion analysis.

**1.2Project Overview**

The project aims to design a web traffic tracking and analysis system to provide insights into user behavior on e-commerce websites. This system will help merchants understand both the general performance of their website and more specific user interactions, helping to shape marketing and engagement strategies.

The key objectives of the system are:

* **Basic tracking**: Capture information like browsers, operating systems, page visits, and referrer sites.
* **Advanced analysis**: Enable path analysis, visitor clustering, conversion rate tracking, and referrer effectiveness.
* **Data extraction and modeling**: Efficiently collect user session data to reveal navigation patterns and use a path tree model to visualize this data.

**1.3 SPECIFICATION**

**1.3.1 Hardware Specifications**

The hardware specifications outline the minimum requirements for the server, database server, and client devices that will be used to run the traffic tracking model.

* **Server**: The server will be used to host the MERN website and run the traffic tracking model. The specifications include:
  + **Processor:** Intel Core i5 or equivalent, which provides a good balance between performance and power consumption.
  + **RAM:** 8 GB or more, which ensures that the server can handle multiple requests and tasks simultaneously.
  + **Storage:** 256 GB or more SSD, which provides fast storage and retrieval of data.
  + **Operating System:** Windows 10, Ubuntu 20.04 or equivalent, which is a popular and stable operating system for servers.
* **Database Server**: The database server will be used to store and manage the data collected by the traffic tracking model. The specifications include:
  + **Processor**: Intel Core i3 or equivalent, which provides a good balance between performance and power consumption.
  + **RAM:** 4 GB or more, which ensures that the database server can handle multiple requests and tasks simultaneously.
  + **Storage**: 128 GB or more SSD, which provides fast storage and retrieval of data.
  + **Operating System**: Windows 10, Ubuntu 20.04 or equivalent, which is a popular and stable operating system for servers.
* **Client Devices:** The client devices will be used to access the MERN website and interact with the traffic tracking model. The specifications include:
  + **Desktop/Laptop:** Intel Core i3 or equivalent, 4 GB RAM, 128 GB SSD, Windows 10 or macOS High Sierra or equivalent, which provides a good balance between performance and power consumption.
  + **Mobile Devices**: Android 10 or iOS 14 or equivalent, which provides a good balance between performance and power consumption.

**1.3.2 Software Specifications**

The software specifications outline the minimum requirements for the frontend, backend, database, and development tools that will be used to build and run the traffic tracking model.

* **Frontend:**
  + **React.js**: version 17 or later, which is a popular and widely-used JavaScript library for building user interfaces.
  + **Redux:** version 7 or later, which is a popular and widely-used state management library for React.
  + **React Router:** version 5 or later, which is a popular and widely-used library for managing client-side routing in React.
  + **CSS:** CSS3 or later, which is a widely-used styling language for web development.
  + **JavaScript:** ECMAScript 6 or later, which is a widely-used programming language for web development.
* **Backend:**
  + **Node.js:** version 14 or later, which is a popular and widely-used JavaScript runtime environment for server-side development.
  + **Express.js:** version 4 or later, which is a popular and widely-used web framework for Node.js.
  + **MongoDB:** version 4 or later, which is a popular and widely-used NoSQL database for storing and managing data.
  + **Mongoose**: version 5 or later, which is a popular and widely-used ORM (Object Relational Mapping) library for MongoDB.
* **Database:**
  + **MongoDB:** version 4 or later, which is a popular and widely-used NoSQL database for storing and managing data.
  + **MongoDB Compass:** version 1 or later, which is a popular and widely-used GUI (Graphical User Interface) tool for managing MongoDB databases.
* **Development Tools:**
  + **Visual Studio Code:** version 1 or later, which is a popular and widely-used code editor for web development.
  + **Git:** version 2 or later, which is a popular and widely-used version control system for web development.
  + **npm:** version 6 or later, which is a popular and widely-used package manager for Node.js.
* **Testing Tools:**
  + **Jest:** version 26 or later, which is a popular and widely-used testing framework for React.
  + **Enzyme:** version 3 or later, which is a popular and widely-used testing library for React.
  + **Cypress:** version 4 or later, which is a popular and widely-used testing framework for web development.

**2. LITERATURE SURVEY**

**2.1 Existing System**

Here are a few more systems similar to Google Analytics:

**1. Matomo (formerly Piwik)**

Matomo is an open-source web analytics platform that provides detailed insights into website traffic, user behavior, and conversion rates. It offers features like heatmaps, session recordings, and A/B testing, making it a popular alternative to Google Analytics.

**2. Adobe Analytics**

Adobe Analytics is a powerful web analytics platform that provides advanced features like predictive analytics, customer journey mapping, and attribution modeling. It's a popular choice among enterprises and e-commerce businesses.

**3. Microsoft Clarity**

Microsoft Clarity is a web analytics platform that provides insights into user behavior, session recordings, and heatmaps. It's a relatively new player in the market, but it's gaining popularity due to its ease of use and integration with Microsoft's other products.

**4. Clicky**

Clicky is a web analytics platform that provides real-time insights into website traffic, user behavior, and conversion rates. It's known for its ease of use and offers features like heatmaps, session recordings, and alerts.

**5. Mixpanel**

Mixpanel is a product analytics platform that provides insights into user behavior, retention, and conversion rates. It's popular among SaaS businesses and e-commerce companies that want to track user engagement and optimize their products.

**6. Chartbeat**

Chartbeat is a web analytics platform that provides real-time insights into website traffic, user behavior, and engagement metrics. It's popular among news organizations and media companies that want to track their online presence.

**7. Crazy Egg**

Crazy Egg is a web analytics platform that provides insights into user behavior, heatmaps, and A/B testing. It's popular among e-commerce businesses and marketers who want to optimize their website's user experience.

**8. Hotjar**

Hotjar is a web analytics platform that provides insights into user behavior, heatmaps, and session recordings. It's popular among e-commerce businesses and marketers who want to understand their website's user experience.

**9. Google Analytics**

Google Analytics is a web analytics service offered by Google that tracks and reports website traffic, user behavior, and conversion rates. It provides insights into how users interact with a website, enabling data-driven decisions to optimize the user experience, improve conversion rates, and increase revenue.

**Problems with existing problem:**

1. **Data Inaccuracy**: Existing systems like Google Analytics use data sampling, which can lead to inaccuracies in reporting.
2. **Data Retention Limits**: Many systems have data retention limits, which can affect the accuracy of historical data.
3. **Cookie Tracking Issues**: Systems that use cookies to track user behavior can be affected by cookie blocking and deletion.
4. **Complexity**: Many systems are complex and require technical expertise to set up and use.
5. **Cost**: Some systems can be expensive, especially for large enterprises.
6. **Limited Customization**: Some systems have limited customization options, making it difficult to tailor the system to specific business needs.
7. **Integration Issues**: Integrating existing systems with other tools and platforms can be challenging.
8. **Security Concerns**: Some systems may have security concerns, such as data breaches or unauthorized access.
9. **Limited Scalability**: Some systems may not be able to handle large volumes of data or traffic.
10. **Limited Support**: Some systems may have limited support options, making it difficult to get help when needed.

**2.2 Proposed System**

The proposed system aims to overcome the limitations of existing traffic tracking and analysis models by introducing a more advanced, site-centric web analytics tool. This system will focus on identifying patterns and regularities in the way users access and use web resources. Instead of relying solely on basic session data, it will incorporate advanced analysis, such as path analysis, clustering, and conversion rate tracking, to provide deeper insights into user behavior and purchasing habits.

A key feature of the proposed system is the use of full-path user session modeling, where each user's navigation is represented as a sequence of actions or page views. This allows for detailed path analysis and the identification of challenges users face while navigating the site. The system will also utilize clickstream data to analyze consumer behavior and measure the effectiveness of marketing actions, while addressing the limitations of traditional clickstream analysis by offering more sophisticated user behavior tracking and segmentation.

Additionally, the system will provide web personalization through the analysis of both sequential and non-sequential user behavior patterns. By utilizing clustering techniques and path tree structures, the system can deliver more customized user experiences, optimize product recommendations, and increase customer satisfaction.

It is also a personalized system for a each website which eliminate the third party allowance. Secondary, We trying to focus on making analysis dashboard for both customer & admin .

**2.3 Feasibility Study**

**I. Executive Summary**

The Traffic Tracking and Analysis Model for the Effective Management of E-commerce Transactions aims to provide a comprehensive model for tracking and analyzing online user behavior, enabling e-commerce businesses to make informed decisions and improve their marketing strategies. This feasibility study assesses the project's viability and provides recommendations for its implementation.

**II. Project Overview**

* **Project Description:** The project involves the development of a traffic tracking and analysis model that can track online user behavior, including page views, clicks, and navigation patterns. The model will provide insights into user behavior, enabling e-commerce businesses to optimize their websites, improve user experience, and increase conversions.
* **Project Location:** The project will be implemented on online e-commerce platforms, including websites and mobile applications.
* **Project Timeline:** The project is expected to be completed within 3 months, depending on the complexity of the implementation.

**III. Market Analysis**

* **Market Demand:** The demand for traffic tracking and analysis tools is high, with many e-commerce businesses seeking to improve their online presence and user experience.
* **Market Size:** The market size for traffic tracking and analysis tools is significant, with many players operating in the market.
* **Competitor Analysis:** The market is competitive, with many established players offering similar solutions. However, the proposed model's unique features, such as its ability to track user behavior across multiple devices and platforms, will differentiate it from existing solutions.

**IV. Technical Feasibility**

* **Technical Requirements:** The project requires the development of a robust and scalable traffic tracking and analysis model that can handle large volumes of data. The model will need to be integrated with existing e-commerce platforms and will require significant computational resources.
* **Technical Risks:** The project is technically feasible, but there are risks associated with data privacy and security. The model will need to be designed to ensure that user data is collected and stored securely.
* **Technical Expertise:** The project requires significant technical expertise, including experience in data analytics, machine learning, and software development.

**V. Financial Feasibility**

* **Capital Costs:** The project requires significant capital investment, including costs associated with software development, hardware, and personnel.
* **Operating Costs:** The project's operating costs will be significant, including costs associated with data storage, computational resources, and personnel.
* **Revenue Projections:** The project's revenue projections are significant, with potential revenue streams including subscription fees, advertising, and data analytics services.
* **Break-Even Analysis:** The project's break-even analysis indicates that it will take approximately 2-3 years to break even, depending on the revenue streams and costs.

**VI. Environmental and Social Impact**

* **Environmental Impact:** The project's environmental impact is minimal, as it does not require significant physical resources or infrastructure.
* **Social Impact:** The project's social impact is significant, as it will enable e-commerce businesses to improve their online presence and user experience, potentially leading to increased sales and revenue.

**VII. Conclusion**

Based on the analysis, the Traffic Tracking and Analysis Model for the Effective Management of E-commerce is feasible and has significant potential for revenue growth. The project's technical feasibility, market demand, and financial viability make it an attractive opportunity for e-commerce businesses.

**VIII. Recommendations**

* **Implementation Plan:** The project should be implemented in phases, with the first phase focusing on the development of the traffic tracking and analysis model.
* **Resource Allocation:** The project requires significant resource allocation, including personnel, hardware, and software.
* **Monitoring and Evaluation:** The project's progress should be monitored and evaluated regularly to ensure that it is meeting its objectives and to identify areas for improvement.

**3. SYSTEM ANALYSIS & DESIGN**

**3.1 Requirement Specification**

**System Name:** Traffic Tracking and Analysis Model for the Effective Management of E-commerce.

**System Overview:**

The Traffic Tracking and Analysis Model is a software system designed to track and analyze online user behavior on e-commerce platforms. The system aims to provide insights into user behavior, enabling e-commerce businesses to optimize their websites, improve user experience, and increase conversions.

**Functional Requirements:**

1. **User Tracking:**
   * The system shall be able to track online user behavior, including page views, clicks, and navigation patterns.
   * The system shall be able to track user behavior across multiple devices and platforms.
2. **Data Collection:**
   * The system shall be able to collect data on user behavior, including demographic information, browsing history, and search queries.
   * The system shall be able to collect data from multiple sources, including websites, mobile applications, and social media platforms.
3. **Data Analysis:**
   * The system shall be able to analyze data on user behavior, including identifying trends, patterns, and correlations.
   * The system shall be able to provide insights into user behavior, including recommendations for improving user experience and increasing conversions.
4. **Reporting and Visualization:**
   * The system shall be able to generate reports on user behavior, including dashboards, charts, and tables.
   * The system shall be able to provide visualization tools, including heat maps, click-through rates, and conversion rates.
5. **Integration:**
   * The system shall be able to integrate with existing e-commerce platforms, including websites and mobile applications.
   * The system shall be able to integrate with third-party services, including social media platforms and advertising networks.

**Non-Functional Requirements:**

1. **Performance:**
   * The system shall be able to handle large volumes of data, including millions of user interactions per day.
   * The system shall be able to provide real-time analytics and insights.
2. **Security:**
   * The system shall be able to ensure the security and integrity of user data, including demographic information and browsing history.
   * The system shall be able to comply with relevant data protection regulations, including GDPR and CCPA.
3. **Scalability:**
   * The system shall be able to scale to meet the needs of large e-commerce businesses, including handling millions of user interactions per day.
   * The system shall be able to adapt to changing user behavior and market trends.
4. **Usability:**
   * The system shall be easy to use, including intuitive interfaces and clear documentation.
   * The system shall be able to provide training and support to users, including online tutorials and customer support.

**Interface Requirements:**

1. **User Interface:**
   * The system shall provide a user-friendly interface, including dashboards, charts, and tables.
   * The system shall provide visualization tools, including heat maps, click-through rates, and conversion rates.
2. **API Interface:**
   * The system shall provide an API interface, including RESTful APIs and SDKs.
   * The system shall be able to integrate with third-party services, including social media platforms and advertising networks.

**Data Requirements:**

1. **Data Sources:**
   * The system shall be able to collect data from multiple sources, including websites, mobile applications, and social media platforms.
   * The system shall be able to collect data on user behavior, including demographic information, browsing history, and search queries.
2. **Data Storage:**
   * The system shall be able to store large volumes of data, including millions of user interactions per day.
   * The system shall be able to ensure the security and integrity of user data, including demographic information and browsing history.

**Testing Requirements:**

1. **Unit Testing:**
   * The system shall be tested at the unit level, including testing individual components and modules.
   * The system shall be tested for functionality, performance, and security.
2. **Integration Testing:**
   * The system shall be tested at the integration level, including testing interactions between components and modules.
   * The system shall be tested for functionality, performance, and security.
3. **System Testing:**
   * The system shall be tested at the system level, including testing the entire system.
   * The system shall be tested for functionality, performance, and security.

**Deployment Requirements:**

1. **Deployment Environment:**
   * The system shall be deployed in a cloud-based environment, including AWS or Azure.
   * The system shall be deployed in a scalable and secure environment.
2. **Deployment Process:**
   * The system shall be deployed using a continuous integration and continuous deployment (CI/CD) process.
   * The system shall be deployed with minimal downtime and disruption to users.

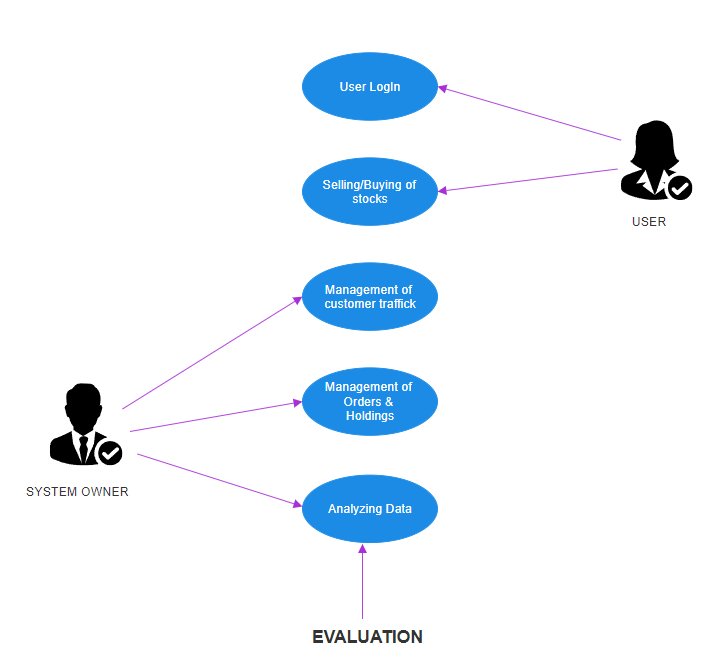
**3.2 Flowcharts / DFDs / ERDs**

This section explains the use of various diagrams to illustrate the flow of data, processes, and user interactions in the system. It includes:

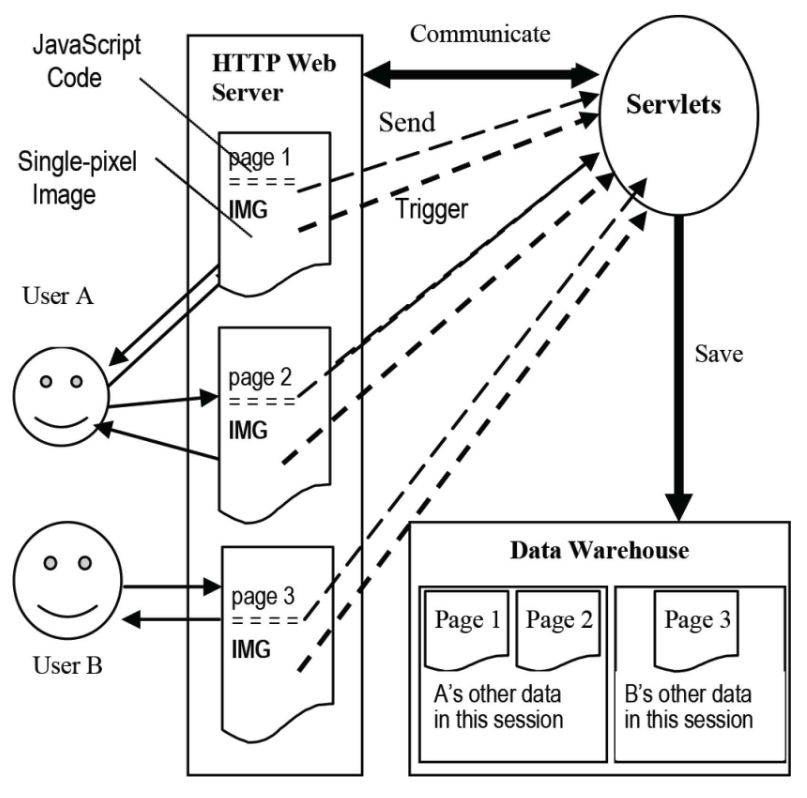
**3.2.1 Use-Case Diagram**

A **Use-Case Diagram** represents the interactions between the system and its users. The diagram shows how different users (actors) interact with the system to achieve various goals. It helps visualize the system’s behavior from an external point of view.

* **Purpose**: To provide a high-level view of the system, focusing on user interactions.
* **Elements**: Actors (users), use cases (tasks or functions), and the relationships between them.



Use-case diagram 1



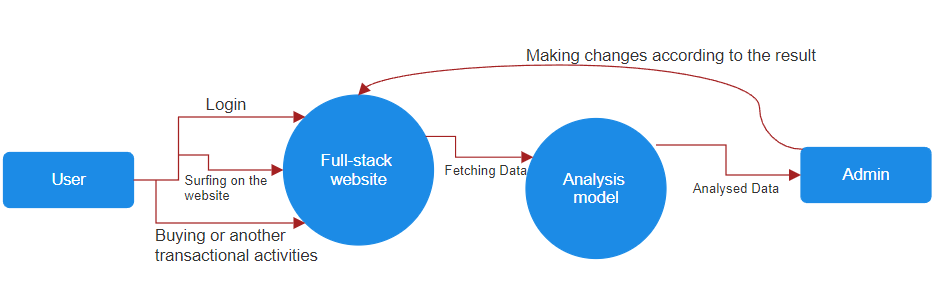
Use-case diagram 2

**3.2.2 Data Flow Diagram (DFD)**

The **Data Flow Diagram (DFD)** is a system model that illustrates how data moves through the system. It is useful for visualizing processes, external entities, and data storage within the system.

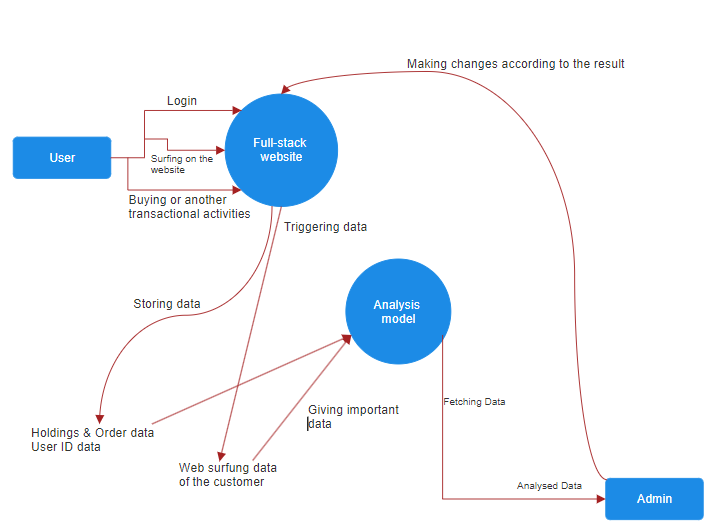
**DFD Level 0**

* **Overview**: This is the top-level DFD, providing a general view of the system as a whole. It captures the major processes and the flow of data between the external entities and the system.



**DFD Level 1**

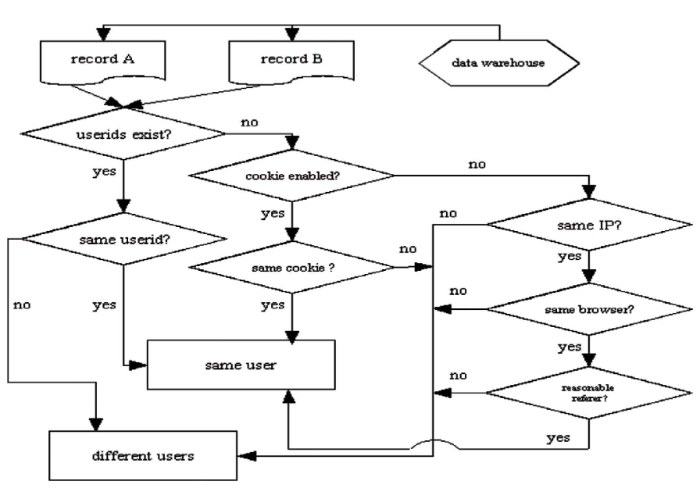
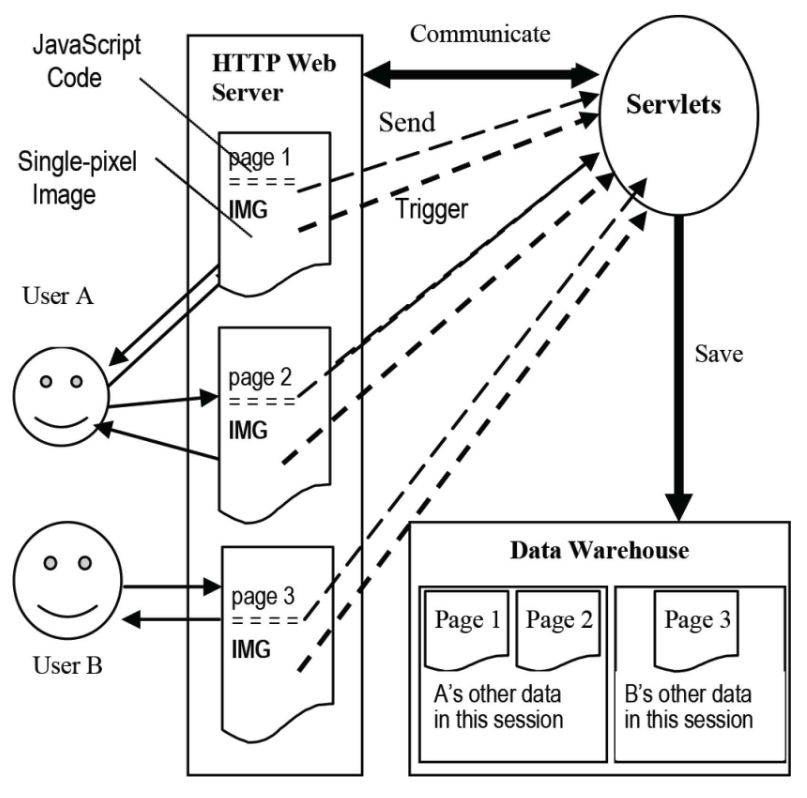
* **Details**: This diagram provides a more detailed breakdown of the main processes from DFD Level 0. It shows how data moves between the sub-processes within the system and includes more granularity in the representation of data stores and interactions.



**3.2.3 Traffic Tracking Diagram**

This is an **inspirational diagram** designed to show the flow of data and processes involved in tracking user behavior on a website. Using **JavaScript tracking** methods, this diagram shows how the model captures data such as user navigation paths, clicks, session data, and other user interactions. This data is critical for business growth and informs key decisions regarding customer engagement and marketing strategies.

* **Purpose**: To demonstrate how the traffic tracking system collects and processes data in real time.
* **Focus**: Building a comprehensive tracking system to capture important information for business owners.



**3.3 Design and Test Steps / Criteria**

Here are the design and test steps for the Traffic Tracking and Analysis Model for the Effective Management of E-commerce Transactions:

**Design Steps:**

1. **Requirements Gathering:**
   * Identify the requirements of the system, including functional and non-functional requirements.
   * Gather information from stakeholders, including e-commerce businesses and users.
2. **System Architecture:**
   * Design the overall architecture of the system, including the components and modules.
   * Identify the technologies and tools to be used, including programming languages, databases, and frameworks.
3. **Component Design:**
   * Design the individual components of the system, including the user tracking module, data collection module, data analysis module, and reporting module.
   * Identify the interfaces and APIs to be used for communication between components.
4. **Database Design:**
   * Design the database schema, including the tables and relationships.
   * Identify the data types and formats to be used for storing and retrieving data.
5. **User Interface Design:**
   * Design the user interface, including the layout, navigation, and visual elements.
   * Identify the user experience and usability requirements.

**Test Steps:**

1. **Unit Testing:**
   * Test individual components and modules, including the user tracking module, data collection module, data analysis module, and reporting module.
   * Test for functionality, performance, and security.
2. **Integration Testing:**
   * Test the interactions between components and modules, including the user tracking module, data collection module, data analysis module, and reporting module.
   * Test for functionality, performance, and security.
3. **System Testing:**
   * Test the entire system, including all components and modules.
   * Test for functionality, performance, and security.
4. **User Acceptance Testing (UAT):**
   * Test the system with real users, including e-commerce businesses and users.
   * Test for usability, user experience, and functionality.
5. **Performance Testing:**
   * Test the system's performance, including response time, throughput, and scalability.
   * Test for performance under different loads and scenarios.
6. **Security Testing:**
   * Test the system's security, including authentication, authorization, and data encryption.
   * Test for vulnerabilities and weaknesses.
7. **Deployment Testing:**
   * Test the system's deployment, including the deployment process and environment.
   * Test for deployment success and system availability.

**Test Cases:**

1. **User Tracking Module:**
   * Test case 1: User tracking module successfully tracks user behavior.
   * Test case 2: User tracking module fails to track user behavior.
2. **Data Collection Module:**
   * Test case 1: Data collection module successfully collects data from multiple sources.
   * Test case 2: Data collection module fails to collect data from multiple sources.
3. **Data Analysis Module:**
   * Test case 1: Data analysis module successfully analyzes data and provides insights.
   * Test case 2: Data analysis module fails to analyze data and provide insights.
4. **Reporting Module:**
   * Test case 1: Reporting module successfully generates reports and visualizations.
   * Test case 2: Reporting module fails to generate reports and visualizations.

**Test Data:**

1. **User Data:**
   * Test data 1: User ID, username, password, and demographic information.
   * Test data 2: User behavior data, including page views, clicks, and navigation patterns.
2. **Data Sources:**
   * Test data 1: Data from multiple sources, including websites, mobile applications, and social media platforms.
   * Test data 2: Data from different formats, including CSV, JSON, and XML.

**Test Environment:**

1. **Hardware:**
   * Test environment 1: Server with 8 GB RAM, 2.5 GHz processor, and 1 TB storage.
   * Test environment 2: Client with 4 GB RAM, 1.5 GHz processor, and 500 GB storage.
2. **Software:**
   * Test environment 1: Operating system, including Windows or Linux.
   * Test environment 2: Browser, including Chrome or Firefox.
3. **Network:**
   * Test environment 1: Internet connection with 100 Mbps bandwidth.
   * Test environment 2: Local area network with 1 Gbps bandwidth.

**3.4 Algorithms**

This section will cover the algorithms implemented in the project, focusing on:

**Path Analysis Algorithm**

* **Description:** The Path Analysis Algorithm is used to track and analyze user navigation paths on the e-commerce website. It helps to identify the most common paths taken by users, the paths leading to conversions, and other insights that can inform marketing and sales strategies.
* **Pseudocode:**
  1. Initialize an empty graph to represent the website's structure.
  2. For each user session, extract the sequence of pages visited.
  3. For each page in the sequence, add an edge to the graph representing the transition from the previous page to the current page.
  4. Calculate the frequency of each edge in the graph.
  5. Identify the most common paths by selecting the edges with the highest frequency.
  6. Analyze the paths leading to conversions by identifying the edges that result in a purchase.
* **Implementation:** The Path Analysis Algorithm is implemented using a graph database, which allows for efficient storage and querying of the website's structure and user navigation paths. The algorithm is integrated into the system through a Python script that extracts user session data from the database and updates the graph accordingly.

**Data Clustering Algorithm**

* **Description:** The Data Clustering Algorithm is used to group users based on their behavior and interests. It helps to identify patterns and trends in user behavior that can inform marketing and sales strategies.
* **Pseudocode:**
  1. Initialize an empty set of clusters.
  2. For each user, extract their behavior and interest data.
  3. Calculate the similarity between each user and each cluster using a distance metric (e.g. Euclidean distance).
  4. Assign each user to the cluster with the highest similarity.
  5. Update the cluster centroids based on the new user assignments.
  6. Repeat steps 3-5 until convergence.
* **Implementation:** The Data Clustering Algorithm is implemented using the K-Means clustering algorithm, which is a widely used and efficient method for clustering data. The algorithm is integrated into the system through a Python script that extracts user behavior and interest data from the database and updates the cluster assignments accordingly.

**Conversion Rate Algorithm**

* **Description:** The Conversion Rate Algorithm is used to calculate the purchase conversion rate based on visitor data. It helps to identify the effectiveness of marketing and sales strategies in driving conversions.
* **Pseudocode:**
  1. Initialize an empty set of conversions.
  2. For each user session, extract the sequence of pages visited.
  3. For each page in the sequence, check if the page is a conversion page (e.g. a purchase page).
  4. If the page is a conversion page, add the user session to the set of conversions.
  5. Calculate the conversion rate by dividing the number of conversions by the total number of user sessions.

**3.5 Testing Process (Test Cases)**

This section will outline the various test cases that were developed to validate the functionality and performance of the system:

**Path Analysis Algorithm**

* **Test Case 1:** Test that the algorithm correctly identifies the most common paths taken by users.
  + Input: A set of user sessions with page views.
  + Expected Output: The most common paths taken by users.
  + Test Steps:
    1. Create a set of user sessions with page views.
    2. Run the Path Analysis Algorithm on the user sessions.
    3. Verify that the algorithm correctly identifies the most common paths taken by users.
* **Test Case 2:** Test that the algorithm correctly identifies the paths leading to conversions.
  + Input: A set of user sessions with page views and conversion data.
  + Expected Output: The paths leading to conversions.
  + Test Steps:
    1. Create a set of user sessions with page views and conversion data.
    2. Run the Path Analysis Algorithm on the user sessions.
    3. Verify that the algorithm correctly identifies the paths leading to conversions.
* **Test Case 3:** Test that the algorithm handles edge cases correctly.
  + Input: A set of user sessions with page views and edge cases (e.g. users who only visit one page).
  + Expected Output: The algorithm should handle edge cases correctly.
  + Test Steps:
    1. Create a set of user sessions with page views and edge cases.
    2. Run the Path Analysis Algorithm on the user sessions.
    3. Verify that the algorithm handles edge cases correctly.

**Data Clustering Algorithm**

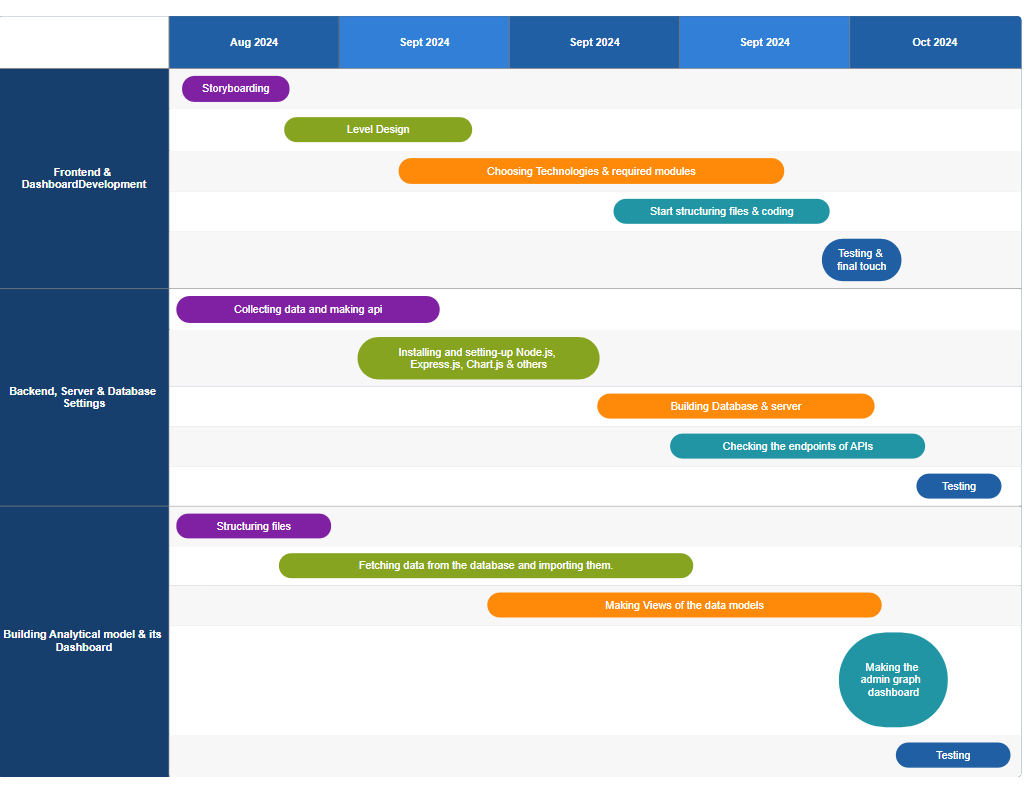
* **Test Case 1:** Test that the algorithm correctly clusters users based on their behavior and interests.
  + Input: A set of user data with behavior and interest information.
  + Expected Output: The algorithm should correctly cluster users based on their behavior and interests.
  + Test Steps:
    1. Create a set of user data with behavior and interest information.
    2. Run the Data Clustering Algorithm on the user data.
    3. Verify that the algorithm correctly clusters users based on their behavior and interests.
* **Test Case 2:** Test that the algorithm handles outliers correctly.
  + Input: A set of user data with outliers (e.g. users with unusual behavior or interests).
  + Expected Output: The algorithm should handle outliers correctly.
  + Test Steps:
    1. Create a set of user data with outliers.
    2. Run the Data Clustering Algorithm on the user data.
    3. Verify that the algorithm handles outliers correctly.
* **Test Case 3:** Test that the algorithm converges correctly.
  + Input: A set of user data with behavior and interest information.
  + Expected Output: The algorithm should converge correctly.
  + Test Steps:
    1. Create a set of user data with behavior and interest information.
    2. Run the Data Clustering Algorithm on the user data.
    3. Verify that the algorithm converges correctly.

**Conversion Rate Algorithm**

* **Test Case 1:** Test that the algorithm correctly calculates the conversion rate.
  + Input: A set of user sessions with conversion data.
  + Expected Output: The algorithm should correctly calculate the conversion rate.
  + Test Steps:
    1. Create a set of user sessions with conversion data.
    2. Run the Conversion Rate Algorithm on the user sessions.
    3. Verify that the algorithm correctly calculates the conversion rate.
* **Test Case 2:** Test that the algorithm handles edge cases correctly.
  + Input: A set of user sessions with edge cases (e.g. users who do not convert).
  + Expected Output: The algorithm should handle edge cases correctly.
  + Test Steps:
    1. Create a set of user sessions with edge cases.
    2. Run the Conversion Rate Algorithm on the user sessions.
    3. Verify that the algorithm handles edge cases correctly.
* **Test Case 3:** Test that the algorithm is robust to changes in the input data.
  + Input: A set of user sessions with conversion data and changes to the input data.
  + Expected Output: The algorithm should be robust to changes in the input data.
  + Test Steps:
    1. Create a set of user sessions with conversion data and changes to the input data.
    2. Run the Conversion Rate Algorithm on the user sessions.
    3. Verify that the algorithm is robust to changes in the input data.

**RESULTS / OUTPUTS**

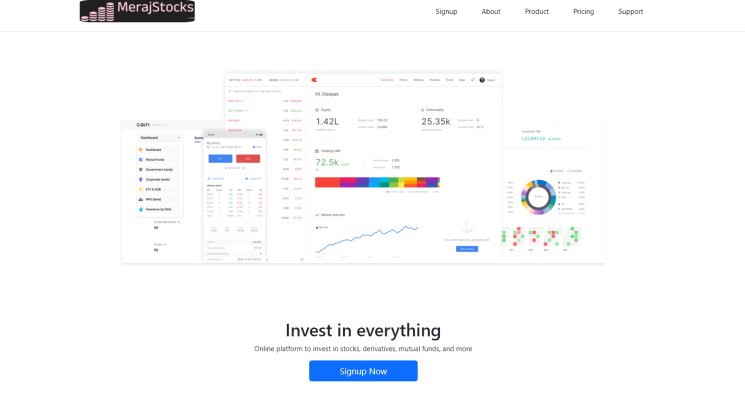
Our implementation environment consists of JavaScript programming language, Node.js, Express.js, React.js and MongoDB & its tools, and a testing site JEST with 12 pages, including both static and dynamic pages. We tested the traffic tracking and analysis tool on a local host. The Gantt-chart (given below) shows architecture website deployment of the tracking and analysis application.



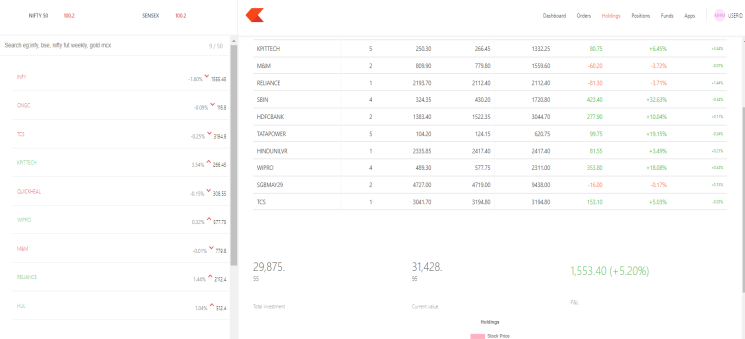
Now, we illustrate how the final result of our project looks like from different aspects of implementation. We also use the snapshots for showing the results of final product. Mainly, it has two view aspects:

**6.1 Customer side view**

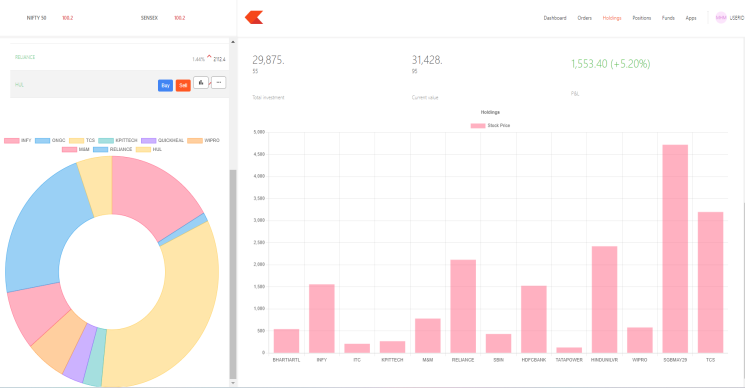
In this section, we use snapshots to show the website structure used by customer for buying and selling which is done based on the given data.



Homepage



Buying-selling interface



Analytics dashboard for customer

**6.2Admin side view**

In this section, we use snapshots to show the admin side dasboard used for tracking & analysing insights which is done based on the given data.





Admin panels

**CONCLUSIONS / FUTURE WORK**

We introduces the design and development of a tracking and analysis model/tool for the effective management of e-commerce. This tool is targeted toward helping e-commerce organizations know more about their users in order to develop efficient marketing strategies.

The contributions of our work include:

The combination of three tracking approaches (i.e., JavaScript approach,Web Server Logs and Client-side server script) allows a user to be tracked always by at least one approach.

The improved client-side script approach has more distinctive advantages over the general web server logs, which requires JavaScript support and server deployment in the user’s browser. The improved client-side script approach does not require JavaScript support to function.

Two main features make the JavaScript approach more effective and different. First, it can track most keyboard and mouse events of online users. The user of this tool can configure what kinds of events about a user should be tracked. Second, it is capable of tracking any form input on the page.

The improved Web server log and JavaScript approaches are designed to work on both static and dynamic Web pages.

In addition to the basic analysis, advanced analysis such as path tree model and user clustering are developed based on the tracked data.

This paper is helpful for designers and developers of e-commerce applications because it highlights some key fundamental and/or pragmatic challenges in the development of e-commerce traffic tracking applications. Therefore, examining the design process, operational efficiency and effectiveness is clearly important because we are able to understand, apply, and promote notable design principles, processes, and practices to overcome the identified challenges.

This study of the design issues in e-commerce traffic tracking systems development deals with issues that ultimately affect the overall system quality and hence the factors that affect customer satisfaction of e-commerce website. Thus, this paper has both academic and practical implications for e-commerce website management, design, and performance enhancement.

In particular, we focused on three broad main challenge areas, namely: (a) how technologies, frameworks, and architectures influence and cause e-commerce applications developers to take certain premature and undesirable design decisions, (b) the interactivity and complexity of e-commerce applications which requires developers to provide several models rather than a single model for the same problem, and finally (c) impact of existing quality assessment criteria for e-commerce traffic tracking applications design.

We plan to extend this work in several directions by addressing the following:

Use a realistic e-commerce site to test this tool. Although implementation of this tool achieved its expectation, we realize a software can behave very differently from testing results when it is used in a real environment.

Develop recommendation and personalization framework based on path analysis, page content analysis, and cluster analysis. For example, suppose we find from the path analysis that the page 3 (i.e., *p*3) usually follows a path *p*1*p*2. We can recommend, therefore, page 3 to a user when the user has finished browsing *p*1*p*2. Another way to do personalization is to combine path analysis with user clusters. We can recommend user B’s path to user A if A and B are in the same group.

Extend the proxy server to handle other protocols, such as https.Web traffic analysis tools can enhance a Web site’s efficiency through the analysis of the users’ data, although more and more people are concerned about privacy issues today. To address privacy concerns in a user friendly way, the users should know what is being tracked about them. Also, the tracking process should not interrupt the users’ normal navigation so the tracking is done in the background. To keep the tracking open to the users, the Web server should be configured to communicate with the users’ browsers. A user’s browser can be configured to request a Web site’s privacy policy, store the user’s privacy preferences and check these preferences against the site policy before any processing. If the Web site’s privacy policy does not match those preferred by the user, the user will be warned or advised. In this way, the user has the choice whether he/she can be tracked, what data can be tracked and how the tracked data can be kept and used.

We are looking forward a model that fulfills our thinkings ,i.e., the proliferation of new data sets and introduction of massive data migration capabilities disrupts existing business models and ecosystems and businesses must embrace data analytics and statistical reasoning to make decisions that improve efficiency, risk management and profits. From using granular data to personalize products and services to scaling digital platforms to match buyers and sellers, businesses use business analytics to enable and enhance better, faster and facts based decision making.

REFERENCES

[[1]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B1)SA Ehikioya and S Lu, A path analysis model for effective e-commerce transactions, Afr. J. Comput. ICT, Vol. 12, 2019, pp. 55-71.

[[2]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B2)“Web Analytics”, Tutorials Point (I) Pvt. Ltd., 2015. Available from: http://www.tutorialspoint.com/web\_analytics\_turorial.pdf (accessed December 12, 2018).

[3]G Zheng and S Peltsverger, Web analytics overview, M Khosrow-Pour (editor), Encyclopedia of Information Science and Technology, third ed., IGI, Global, Hershey, PA, USA, 2015.

[[4]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B4)[MT Nguyen, TD Diep, T Hoang Vinh, T Nakajima, and N Thoai, Analyzing and visualizing web server access log file, T Dang, J Küng, R Wagner, N Thoai, and M Takizawa (editors), Springer, in International Conference on Future Data and Security Engineering (FDSE 2018) — Lecture Notes in Computer Science, vol 11251 (Cham, 2018), pp. 349-367.](https://doi.org/10.1007/978-3-030-03192-3_27)

[[5]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B5)B Dykes, Web analytics kick start guide: a primer on the fundamentals of digital analytics, Adobe Press Books, Peachpit Press, Pearson Education, San Francisco, CA, USA, 2014.

[[6]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B6)A Kaushik, Web analytics 2.0: the art of online accountability and science of customer centricity, John Wiley & Sons, Hoboken, NJ, USA, 2010.

[[7]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B7)ET Peterson, Web analytics demystified: a marketer’s guide to understanding how your web site affects your business, Celilo Group Media and CafePress, Portland, OR, USA, 2004.

[[8]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B8)B Clifton, Advanced web metrics with Google analytics, third ed., Wiley Publishing, Hoboken, NJ, USA, 2012.

[9]A Kaushik, Web Analytics: An Hour a Day, Wiley Publishing, Hoboken, NJ, USA, 2007.

[10]A Croll and S Power, Complete web monitoring: watching your visitors, performance, communities, and competitors, O’Reilly Media Inc., Sebastopol, CA, USA, 2009.

[[11]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B11)S Jackson, Cult of analytics: driving online marketing strategies using web analytics, Butterworth-Heinemann, Oxford, United Kingdom, 2009.

[[12]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B12)I Bekavac and DG Praničević, Web analytics tools and web metrics tools: an overview and comparative analysis, Croat. Oper. Res. Rev., Vol. 6, 2015, pp. 373-386.

[13]BJ Jansen, Understanding user-web interactions via web analytics, Morgan & Claypool, Williston, VT, USA, 2009.

[[14]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B14)[D Booth and BJ Jansen, A review of methodologies for analyzing websites, BJ Jansen, A Spink, and I Taksa (editors), Handbook of Research on Web Log Analysis, IGI Global, Hershey, PA, USA, 2010, pp. 141-162.](https://doi.org/10.4018/978-1-59904-974-8.ch008)

[[15]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B15)[RE Bucklin and C Sismeiro, Click here for internet insight: advances in clickstream data analysis in marketing, J. Interact. Market., Vol. 23, 2009, pp. 35-48.](https://doi.org/10.1016/j.intmar.2008.10.004)

[[16]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B16)A Ezzedin, Tracking product journey from carting to purchasing: 15 secrets to perfecting your online store, E-Nor Inc., 2014. Available from: https://www.e-nor.com/wp-content/uploads/pubs/ebooks/tracking-product-journey-from-carting-to-purchasing.pdf.

[[17]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B17)Eric Fettman, Google analytics universal guide: best practices for implementation and reporting, E-Nor Inc., 2014. Available from: <https://www.e-nor.com/blog/ebooks/google-analytics-universal-guide-best-practices-for-implementation-and-reporting>.

[[18]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B18)[J Andersen, A Giversen, AH Jensen, RS Larsen, TB Pedersen, and J Skyt, Analyzing clickstreams using subsessions, ACM, in Proceedings of the ACM Third International Workshop on Data Warehousing and OLAP (DOLAP) (Washington DC, USA, 2000), pp. 25-32.](https://doi.org/10.1145/355068.355312)

[[19]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B19)L Clark, IH Ting, C Kimble, PC Wright, and D Kudenko, Combining ethnographic and clickstream data to identify user web browsing strategies, Inform. Res., Vol. 11, 2006. Available from: <http://informationr.net/ir/11-2/paper249>.

[[20]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B20)[HK Ellonen, P Wikstrom, and A Johansson, The role of the website in a magazine business: revisiting old truths, J. Media Bus. Stud., Vol. 12, 2015, pp. 238-249.](https://doi.org/10.1080/16522354.2015.1107334)

[[21]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B21)JP de Almeida Ribeiro, The use of web analytics on a small data set in an online media company: shifter’s case study, NOVA Information Management School, Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, 2016. Master’s Degree in Information Management

[[22]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B22)M Lindén, Path analysis of online users using clickstream data: case online magazine website, LUT School of Business and Management, Lappeenranta University of Technology, 2016. Master’s Degree in Strategy, Innovation and Sustainability

[[23]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B23)R Kumar Jain, RS Kasana, and S Jain, Efficient web log mining using doubly linked tree, Int. J. Comput. Sci. Inform. Secur., Vol. 3, 2009, pp. 5.

[[24]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B24)SK Pani, L Panigrahy, VH Sankar, BK Ratha, AK Mandal, and SK Padhi, Web usage mining: a survey on pattern extraction from web logs, Int. J. Instrum. Control Autom., Vol. 1, 2011, pp. 15-23.

[[25]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B25)AT Siddiqui and S Aljahdali, Web mining techniques in e-commerce applications, Int. J. Comput. Appl., Vol. 69, 2013, pp. 39-43.

[[26]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B26)N Jokar, AR Honarvar, S Aghamirzadeh, and K Esfandiari, Web mining and web usage mining techniques, Bulletin Royale Soc. Sci. Liège, Vol. 85, 2016, pp. 321-328.

[[27]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B27)M Wang, G Liu, C Yan, and C Jiang, Modeling and vulnerable points analysis for e-commerce transaction system with a known attack, Springer, in Ninth International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (SpaCCS) (Cham, 2016), pp. 422-436. Lecture Notes in Computer Science, vol. 10066,

[[28]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B28)[Z Tian, Z Zhan, and X Guan, A new structural analysis model for e-commerce ecosystem network, Int. J. Hybrid Inform. Technol., Vol. 7, 2014, pp. 43-56.](http://dx.doi.org/10.14257/ijhit.2014.7.1.04)

[[29]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B29)[MW Lewis and BJ White, SOLO: a linear ordering approach to path analysis of web site traffic, INFOR Inf. Syst. Oper. Res., Vol. 50, 2012, pp. 186-194.](https://doi.org/10.3138/infor.50.4.186)

[[30]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B30)[KN Asha and R Rajkumar, Survey on web mining techniques and challenges of e-commerce in online social networks, Indian J. Sci. Technol, Vol. 9, 2016, pp. 1-5.](https://doi.org/10.17485/ijst/2016/v9i13/85481)

[[31]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B31)[AL Montgomery, S Li, K Srinivasan, and JC Liechty, Modeling online browsing and path analysis using clickstream data, INFORMS, Vol. 23, 2004, pp. 469-631.](https://doi.org/10.1287/mksc.1040.0073)

[[32]](https://www.atlantis-press.com/journals/ijndc/125940874/view#first-bibr-B32)A Noreika and S Drąsutis, Website activity analysis model, Inform. Technol. Control, Vol. 36, 2007, pp. 268-272.