team contract 9/6
project proposal 10/11
fall design poster 11/15
fall poster session 12/2
preliminary design report 12/9

VCU College of Engineering

CS 25-322 AI-generated planning insights powered by Clickstream data

Preliminary Design Report

Prepared for

Mahesh Nair / Tyler Jordan / Emily Croxall

Capital One

By

Priya Choudhary, Bindi Patel, Carissa Trieu, Ivan Emdee

Under the supervision of

Thomas Gyeera

12/9/2024

Executive Summary

The purpose of this project is to collect and analyze real-time clickstream data to provide AI-driven insights for optimizing roadmap planning. Analysis of user interactions on a web application similar to the Capital One Help Center page will allow stakeholders to make better informed decisions with an emphasis on metrics like Average Handle Time and user engagement.

The primary objectives and deliverables of the project include the creation of an intuitive user interface, analysis of clickstream, and implementation of AI/ML models to offer actionable recommendations. A scalable database will also be utilized to handle clickstream data. Using the recommendations, epics, and stories will be automatically generated in JIRA for efficient decision-making.

The final project deliverable will be a fully functional prototype, ready for demonstration to Capital One. Academic deliverables include a project proposal, a preliminary design report, a fall semester poster and presentation by November 2024, a final design report, and a final EXPO poster/ presentation by Spring 2025. This project is relevant for Capital One's Agent Servicing team, as it will allow stakeholders to gain a better understanding of user behavior, improve platform efficiency, and support long-term strategic planning. Ultimately, this project aims to both enhance user experience and operational efficiency.

Table of Contents

Section A. Problem Statement

In today's fast-paced digital world, platform stakeholders need to make informed, data-driven decisions to enhance user experiences, optimize application performance, and plan for future growth. Clickstream data, which captures user interactions within an application, provides valuable insights into user behavior, but many organizations struggle to fully leverage this information to guide their product roadmaps.

Clickstream is a sequence of pages to describe the history of a user's session. This type of data has proven useful in web usage mining and in generating real-time predictions (Bucklin & Sismeiro, 2009). Currently, existing solutions often fall short when it comes to processing clickstream data at scale because it is done by hand. This takes a significant amount of time to do. Delivering AI-powered recommendations, and integrating seamlessly with tools like JIRA for project management will help shorten the time taken. This project aims to bridge that gap by developing a web application that not only captures clickstream data but also uses AI to analyze it and provide actionable recommendations for improving platform features and reducing inefficiencies.

The system will address key stakeholder questions, such as identifying areas to reduce Average Handle Time (AHT), pinpointing where users spend the most time in the application, and highlighting critical features. Additionally, it will integrate with JIRA to automatically generate mock Epics and stories based on accepted recommendations, helping streamline the implementation process. By creating this solution, the project equips stakeholders with a powerful tool to turn clickstream data into meaningful insights, allowing for smarter, data-driven decisions that support strategic roadmap planning.

Our project client is Capital One, they have tasked us with this project because they want to see if it is more efficient to have an AI look over their clickstream data. Rather than having a human look over it. After doing much research, we believe that it will be and hope to prove it with our prototype. The articles we read said that it was faster and more accurate to have an AI look over the data.

The general field of study that the project falls under is data analytics and artificial intelligence, specifically clickstream analysis and artificial intelligence-driven product roadmap planning. By analyzing the sequence of clicks taken by users on a website, clickstream data can provide valuable customer insights and enhanced business intelligence. Other use cases include web movement investigation, statistical surveying, and programming testing (Hanamanthrao & Thejaswini, 2017). Additionally, AI and machine learning models are transforming how product roadmap planning is done. One key advantage of its implementation is the acceleration of timeline creation. Furthermore, AI and ML algorithms have already been implemented in other industries like traffic management and urban planning to make real-time decisions, based on data collected from GPS and sensors (Khare & Arora, 2024).

Capital One, the sponsor company, is a leading company in the banking industry with a great emphasis on information and technology. Specifically, the Agent Servicing team serves to enhance customer support by providing agents with the necessary tools to effectively assist customers. They provide services that seamlessly handle customer inquiries by enabling agents to access relevant real-time information such as customer account details, transaction histories, and service records. Currently, the platform does not have any integration with artificial intelligence.

Clickstream data has been used to generate AI insights for project and product planning, but it is less common in industries related to customer support platforms for large-scale digital services in the banking industry. For instance, historically, clickstream data has been utilized for optimizing websites and predicting user behavior. In a paper titled “Real-Time Clickstream Data Analytics and Visualization”, Hanamanthrao & Thejaswini (2017) highlighted the benefits of real-time data processing over traditional batch processing techniques for clickstream analysis. Data collected from an online learning platform website included course enrollment, navigation patterns, and time spent per page. The data was stored and processed using Apache Hadoop, an open-source big data storage platform. A real-time data pipeline was implemented using Apache Kafka for data streaming, Apache Spark for data processing, and ElasticSearch for data indexing and searching. Real-time processing led to the creation of immediate insights, such as the identification of popular courses and optimal navigation pathways on the portal. The results also demonstrated that real-time processing improved course recommendation accuracy and resource allocation, compared to traditional batch processing. Previous approaches required more time to process data before any insights could be applied. Manipulation of real-time data processing sets the groundwork for future predictive analytics related to improvements based on live user behavior. Furthermore, this approach allows for future implementation of ML models which create personalized recommendations.

While Hanamanthrao & Thejaswini demonstrated the clear effectiveness of real-time processing, there was no implementation of AI or ML algorithms. Gumber et al.conducted a study, “Predicting Customer Behavior by Analyzing Clickstream Data”, which emphasized the importance of clickstream in understanding user behavior and applying it to develop a machine learning model. In this academic journal, they also reference Hanamanthrao & Thejaswini in the literature survey, noting that they were unable to collect data from multiple sources. Clickstream data was collected from an e-commerce site because shoppers often navigate through multiple pages before making a purchase decision. Unstructured data was obtained from Kaggle, an open-source platform providing datasets, containing information such as event time, event type (view, cart, purchase), and category information. Preprocessing was conducted before implementing the machine learning algorithm to make it usable. The machine learning model chosen, XGBoost, can handle large datasets well while processing them quickly with fewer computing resources. It also utilizes regularization techniques to prevent overfitting, a common machine learning behavior that can potentially give inaccurate predictions, more efficiently than other boosting algorithms. Using the Extreme Gradient Boosting (XGBoost) algorithm, customer behavior predictions were made with 85.9% accuracy and 91.04% recall (Gumber et al., 2021).

Though the industry in which the research was performed is vastly different from agent servicing platforms, similar principles could be applied to enhance customer support platforms.

"Neural Networks for Customer Classification Through Clickstream Analysis" written by Erika Severeyn, Alexandra La Cruz, Roberto Matute, and Juan Estrada in 2023 goes over how most organizations currently interpret large amounts of data within their databases and how they hope to move toward AI algorithms in the future. The article uses the term business intelligence (BI), which is the strategies, technologies, and tools that an organization uses to analyze data. They then use that information to try and gain insights to conduct future business decisions. By analyzing this data organizations can attempt to make predictions about customers behavior, market trends, and other business-impacting factors. The newest addition to companies BI's has been the use of AI to break down these large data sets. Leveraging clickstream data is a valuable resource that can significantly aid businesses in enhancing their productivity through the implementation of artificial intelligence for data analysis. The main use of this so far has been to personalize online experiences for customers. The type of AI being used is an AI that has a neural network. They mimic having a human brain to learn as they get fed more information. By using these neural networks businesses can more accurately predict user behaviors. For the research done in the paper, they used Clickstream data from IMOLKO C.A.'s website. IMOLKO C.A. is a service company that helps businesses increase their profits, retain clients, and reduce the churn rate. The dataset was collected by Google Analytics which is a platform that can collect real-team user interactions. The researchers' goal was to see which users would become customers based on the dataset. They trained the neural network on the pre-processed dataset. They did five different training techniques, it was performed for 90% training and 10% testing, 80% training and 20% testing, 70% training and 30% testing, 60% training and 40% testing, and 50% training and 50% testing. The result was that the different techniques did not affect the AIs accuracy. It stayed consistent throughout all the training. The model demonstrated that it could accurately identify positive cases, cases where the user did become a customer, 85% of the time. It was much better at identifying when a user would not become a customer, achieving an accuracy of 91%.

Another article that we looked into was focused less on trying to make a profit and more on helping users correctly use their website. The article is called "A Machine Learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks" and it was written by Esther Ulitzsch, Vincent Ulitzsch, Qiwei He, and Oliver Lüdtke in 2022. They wanted to see if they could get an AI to predict whether or not a user would fail an interactive problem based on their actions on the website. The thought process was that there was sufficient information for predicting outcomes based on which actions they performed early on and how long it took them. They wanted to be able to predict the outcome as early and accurately as possible. This was done to see if there was a pattern in the way a user would attempt to solve a problem and their success rate. If there was, the website could be redesigned to try and raise the percentage of successful problem-solving. They tested users with two different problems in their study. They called these problems "Lamp Return" and "Meeting Rooms". "Lamp return" involves going through an online shop to return a lamp. "Meeting Rooms" involved going through an online website to reserve rooms based on different requests.

The data from 6,791 “Lamp Return” problems and 6,629 “Meeting Rooms” problems were used for analysis. The success rate for both problems was around 50%. Those who failed generally did fewer actions and spent less time on the problems. The data was preprocessed to make it easier for the AI to understand when it was trying to predict outcomes. This researcher also used XGBoost to try and predict the outcome. The AI was trained on the data and it made its predictions based on many formulas set up by the researchers. The results were that the AI was able to achieve excellent classification performance for the “Lamp Returns” problem. The results were not as clear for the “Meeting Rooms” problem. This comes from the AI’s most predictive feature being time elapsed until action. This means how long the user would spend reading the problem before doing anything. It was much more of a telling feature for the “Lamp Returns” problem than it was for the “Meeting Rooms” problem. Overall though the AI was very accurate at predicting the outcome of the test very early on in the testing. The difference between the accuracy in the problems only shows how important the criteria the AI follows for its predictions are.

Section B. Engineering Design Requirements

This section outlines the core goals, objectives, design specifications, and constraints that guide the development of this project. It provides a structured framework for defining the problem space, informed by client needs, and outlines the conditions and limitations within which the design will operate. By clearly establishing these requirements, the project ensures alignment with stakeholder expectations and creates a roadmap for delivering a successful solution.

The engineering design requirements address several critical aspects that guide the development of the project. First, the Project Goals section describes the overarching goals from the client’s perspective, focusing on the high-level outcomes the client expects. These goals do not provide specific details about the design but highlight the purpose and intent behind the project.

Next, the Design objectives focus on specific, measurable, and time-bound goals that define what the design will achieve. These objectives are drawn from the client’s needs and translated into actionable tasks for the design, ensuring that the solution meets the expected functionality.

The Design Specifications and Constraints section lists the measurable and testable limitations that must be met for the design to be considered successful. It includes performance requirements, hardware capabilities, cost constraints, and data handling regulations, ensuring the design adheres to both functional and legal expectations.

B.1 Project Goals (i.e. Client Needs)

The project aims to help Capital One stakeholders make informed, data-driven decisions using clickstream data. This project will address the underutilization of clickstream data in

roadmap planning and will empower stakeholders to gain actionable insights into user behavior. the primary goals of the project are:

- To leverage real-time clickstream data to provide AT-driven insights for platform optimization.
- To build a web application that mimics Capital One's Help Center and allows stakeholders to visualize data in real time.
- To automate the generation of recommendations for improving key metrics like Average Handle Time (AHT) and user engagement.
- To seamlessly integrate the insights and recommendations into JIRA, allowing for the automatic creation of epics or stories that align with AI-driven suggestions.

B.2 Design Objectives

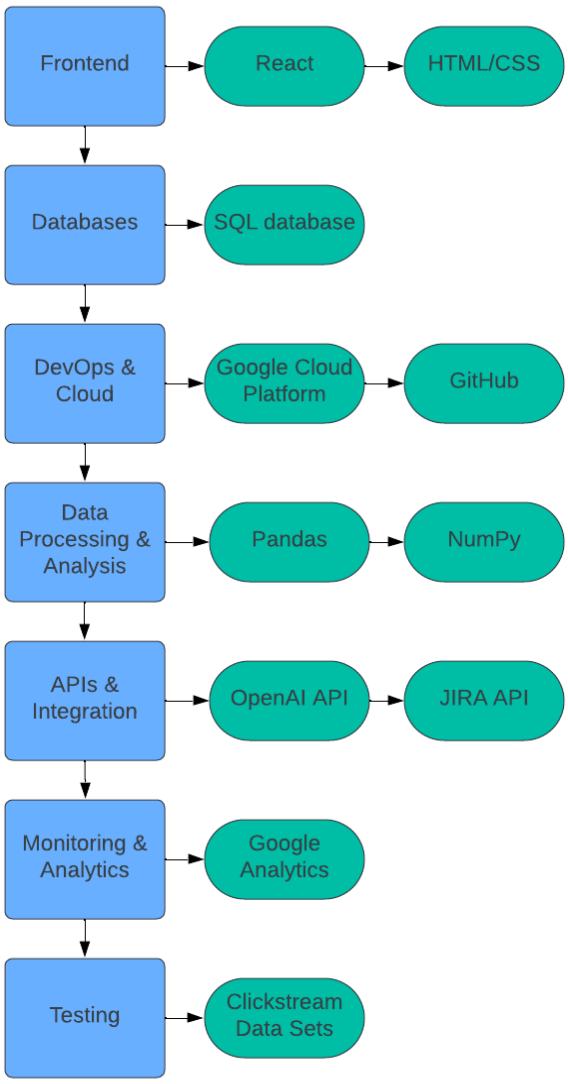
- The design will capture real-time clickstream data from users interacting with a web application that mimics the Capital One Help Center
- The design will provide AI-generated insights based on the clickstream data to answer queries such as, "How can I reduce AHT in the X container?" or "What are my critical features?"
- The design will integrate with JIRA via APIs to allow stakeholders to automatically create epics or stories based on the recommendations provided by the AI model.
- The design will offer a user-friendly, responsive interface that provides easy navigation, filtering, and data visualization using charts and graphs to help users analyze trends and performance metrics.
- The design will ensure that the platform operates effectively across different devices (desktop, mobile, tablet) to accommodate diverse users.

B.3 Design Specifications and Constraints

- The application must handle real-time clickstream data processing with a latency of no more than 5 seconds to ensure accurate insights
- The AI model must analyze clickstream data and generate recommendations with an accuracy rate of at least 85%. the system will utilize OpenAI's GPT for generating insights.
- The system must handle up to 1,000 concurrent users and process data efficiently without performance degradation. This ensures the platform's scalability for larger user bases.
- The application must interface with JIRA via REST APIs to create epics or stories automatically from AI recommendations within 3 seconds of acceptance.
- Clickstream data must be stored in a scalable database capable of handling data growth of at least 1GB/day without affecting system performance.

- The system must comply with Capital One's privacy regulations, ensuring that no personal identifiable information (PII) is exposed or stored. Data encryption will be implemented for both in-transit and at-rest data.
- The UI should allow for filtering of data by various parameters (e.g., date range, container type, user actions) and ensure responsiveness on mobile, tablet, and desktop devices. Performance tests will ensure UI responsiveness within 2 seconds across devices.
- The system should have an uptime of at least 99%, ensuring minimal downtime for continuous access to insights and recommendations.

Tech Stack Diagram:



Section C. Scope of Work

The scope of this project involves building an AI-powered web application to analyze clickstream data for platform performance recommendations, with integration into JIRA for project management. The key objectives include capturing and analyzing real-time data, generating actionable insights, and streamlining roadmap planning. This project is critical for platform stakeholders to make data-driven decisions to improve user engagement and other key metrics.

Boundaries for the project include a focus on web application development with specific deliverables and milestones as outlined below. The team will operate under the Agile methodology, utilizing iterative development to adapt as insights from earlier project phases are incorporated into subsequent stages. The team will work closely with the faculty advisor and project sponsor to ensure timely completion and adherence to the scope, preventing scope creep.

Stakeholder Involvement:

The project sponsor and faculty advisor play a crucial role in verifying and approving each stage of the project. Regular reviews and feedback sessions (weekly Zoom meetings) will be conducted to ensure that the project remains aligned with the expectations and requirements set by the stakeholders. This will help identify any potential issues early on, ensuring any necessary adjustments are made without deviating from the project's scope. Active stakeholder engagement will mitigate the risk of scope creep and help ensure that all deliverables meet the desired quality standards.

C.1 Deliverables

Project Deliverables:

- Web Application: A functioning web app with a user-friendly interface mimicking the Capital One Help Center page. The app will capture and display real-time clickstream data, providing actionable AI-driven insights.
- Clickstream Data Analysis: Real-time data processing and visualization using AI to recommend improvements, including metrics like Average Handle Time (AHT) and user engagement.
- JIRA Integration: Automatic creation of mock Epics and stories based on AI recommendations.
- AI/ML Recommendation Engine: Integration of OpenAI's API for generating insights from clickstream data.
- Database: A scalable SQL database for storing structured clickstream data.
- Final Prototype: A complete, functional prototype that can be demonstrated for review.

Academic Deliverables:

- Team Contract
- Project Proposal
- Preliminary Design Report
- Fall Poster and Presentation (November 2024)
- Final Design Report
- Capstone EXPO Poster and Presentation (Spring 2025)

Risks and Mitigations:

- Access to campus: Some team members may not have consistent access to campus facilities due to living off-campus, but the majority of work can be completed remotely.
- Remote Work: Resources like shared drives, GitHub, and remote collaboration tools (VS Code, JIRA, and Slack) will be used to mitigate risks.
- Third-party vendor delays: The use of existing APIs and software mitigates the need for extended lead times or delays related to third-party ordering.

C.2 Milestones

Milestone	Estimated Completion Date
Team Contract - Establish a foundation for team collaboration and communication. The document lists individual team member strengths, roles, and responsibilities, time commitments, and the communication/meeting structure.	September 6, 2024
Project Proposal - Frame the project early on by outlining key details such as background information, objectives, and the unmet engineering need being addressed. Includes sections on background, literature review, project goals, objectives, constraints, project scope, deliverables, organizational structure, and a proposed timeline.	October 11, 2024

Prototype Web Application Build - Develop a basic web application version to mimic the Capital One Help Center page, integrating key features and functionalities to create a working model for testing and feedback.	November 15, 2024
Fall Design Poster - Present preliminary design results in a public forum at the end of the semester, where team members can showcase their work to sponsors, faculty, administrators, and peers.	November 15, 2024
Preliminary Design Report - Provide a comprehensive summary of the team's design efforts over the semester, with more detail than the project proposal.	December 9, 2024
Data Collection and AI Integration - Implement a system for collecting relevant data, followed by integrating AI/ML models to analyze and interpret the data, providing insights for future development.	February 2025
AI Recommendations in JIRA - Deploy the AI-powered solution, enabling it to analyze clickstream data and provide actionable recommendations. These will be integrated with JIRA to create mock Epics and stories for roadmap planning.	March 2025
Final Design Report Submission - Deliver a detailed report summarizing the team's work, including final design, implementation results, and project outcomes, showcasing the full scope of the project.	April 2025
Capstone Design Expo - Present the final project at a public expo, demonstrating the completed work to industry sponsors, faculty, and	April 24-25, 2025

peers, emphasizing technical achievements and project outcomes.	

C.3 Resources

1. Software:
    a. Integrated Development Environments (IDEs): Primarily using VS Code for development.
    b. Version Control: GitHub will be used to manage source code, track changes, and collaborate.
    c. Data Analysis Platforms: Pandas and NumPy libraries for processing and analyzing clickstream data.
2. Cloud Computing Services:
    a. Amazon Web Services (AWS) or Google Cloud Platform (GCP): To scale real-time data processing and storage as needed, especially during peak usage or testing.
3. APIs and Libraries:
    a. OpenAI API: For generating AI-driven recommendations based on clickstream data.
    b. JIRA API: To automate the creation of stories and epics in JIRA based on AI insights.
4. Databases:
    a. PostgreSQL Database: For storing structured clickstream data for real-time access and analysis.
5. Testing and Analytics Tools:
    a. Clickstream Data Sets: Operational datasets for testing AI recommendations and simulating user interactions on the platform.
    b. Google Analytics: To track and analyze web traffic, providing additional insights into user behavior and augmenting clickstream data for more robust analysis.

At this stage of the project, we do not anticipate purchasing any additional software or resources. However, as the project progresses and requirements evolve, we may reassess our needs and consider acquiring specific tools or services if necessary to ensure successful completion.

Section D. Concept Generation

Several design concepts were explored to address the challenges of utilizing clickstream data for roadmap planning and performance optimization. Each concept was evaluated based on its ability to meet the project's objectives and specifications.

The first design concept is the creation of a modular web application with AI integration. This addresses the design problem by providing a standalone platform for capturing and analyzing clickstream data in real-time, offering stakeholders actionable insights for roadmap planning. The application includes features such as a user-friendly interface for data visualization, AI-driven recommendations to improve metrics like Average Handle Time (AHT) and user engagement, and seamless integration with tools like JIRA to automate the creation of stories and epics based on the AI suggestions. Potential pros include the flexibility to fully customize the design and functionality of the application, enabling it to meet specific stakeholder needs. Additionally, the modular architecture allows for future scalability and integration of new features or tools. Potential cons include the significant development effort required to design a user-friendly interface and implement robust AI and integration features. This could lead to longer development timelines compared to simpler designs.

The second design concept is a backend-focused system that processes clickstream data and delivers insights through external reports or dashboards. This design focuses on building a strong backend that efficiently analyzes data and generates AI-driven recommendations, which are then shared via tools like Tableau or Microsoft Power BI. Potential pros include faster development since the design skips building a full frontend interface and instead relies on existing reporting tools. Potential cons include limited interactivity for users, as they cannot directly query or explore the data in real-time. Additionally, potential risks of failure include challenges in ensuring compatibility with third-party tools and losing control over the customization of how insights are presented.

The third design concept is an AI-powered dashboard that directly displays insights and recommendations. Unlike the second concept, this solution includes a lightweight, custom frontend for users to explore the data visually with options like filtering, sorting, and viewing trends. This design puts a bit more emphasis on the user experience compared to Concept 2. Potential pros include a user-friendly platform that gives stakeholders immediate access to insights in a visual format. Potential cons include less scalability and fewer advanced processing features compared to a backend-focused system. Additionally, potential risks of failure include not meeting all user needs for customization or handling large-scale data efficiently.

The fourth design concept is an AI-assisted query system that lets stakeholders ask natural language questions like "How can I improve AHT?" and receive instant insights. This design is all about simplifying the user experience by providing targeted answers instead of requiring users to interpret dashboards or reports. Potential pros include a highly intuitive and direct way to access insights without navigating complex interfaces. Potential cons include limited visualization options, as users rely on text-based responses rather than charts or graphs.

Additionally, potential risks of failure include challenges in training the AI to accurately interpret natural language queries and deliver actionable results.

Section E. Concept Evaluation and Selection

Each concept was evaluated based on how it was able to meet the project's goals and requirements. The criteria included real-time data processing, user interactivity, scalability, AI functionality, and ease of implementation. A decision matrix was used to compare the strengths and weaknesses of the four concepts.

The first concept, the web application with AI integration, scored highly in flexibility, user interactivity, and scalability. Its ability to provide a comprehensive, standalone solution aligns closely with the project’s goals. However, the development complexity and extended timeline required to create a fully customized interface were considered drawbacks.

The second concept, the backend-focused system with external reporting, excelled in backend efficiency and faster development timelines due to its reliance on existing reporting tools. While it offers in depth AI processing capabilities, the missing frontend component would limit user interactivity and customization. This makes it less suitable for stakeholders who require direct access to insights.

The third concept, the AI-powered dashboard, provides a balanced solution by combining a lightweight frontend with robust AI-driven insights. It allows stakeholders to visualize data and access recommendations in a user-friendly format, making it a strong choice for addressing the project's user experience goals. However, its scalability and data handling capabilities are limited compared to designs with heavier backend development focus.

The fourth concept, the AI-assisted query system, stands out for its simplicity and direct usability. By enabling stakeholders to ask natural language questions and receive instant recommendations, it minimizes the learning curve for users. Despite its intuitive nature, the lack of data visualization and the complexity of natural language processing reduce its feasibility for directly addressing our project objectives.

The web application with AI integration was selected as the final design concept. It provides the most comprehensive solution for capturing, analyzing, and delivering clickstream insights. The architecture allows for scalability and future extensibility. The AI-driven features directly support roadmap planning and performance improvement. Although the development timeline may be longer, the flexibility and alignment with stakeholder needs make it the best choice for this project.

Table 1. Decision Matrix.

	Design Concept 1	Design Concept 2	Design Concept 3	Design Concept 4
Real time processing	9	7	7	5
Scalability	8	9	6	6
Ease of implementation	6	8	7	7
AI functionality	8	7	7	6
UI and Interaction	9	4	8	6
Total Score	8	7	7	6

Section F. Design Methodology

The design methodology for this project follows an iterative engineering design process to ensure that the final system meets all outlined objectives and specifications while addressing the client's needs effectively. The iterative approach involves continuous evaluation, improvement, and validation of the design through computational modeling, experimental testing, and stakeholder feedback. The methodology integrates physical principles of data processing and system optimization, as well as computational tools and testing protocols, to verify that the design meets all performance specifications. Validation ensures that the system functions as intended to provide real-time insights, seamless JIRA integration, and an intuitive user experience.

The project is guided by key principles of modularity, scalability, and user-centric design. These principles inform every phase of development, from architectural planning to final validation. Computational methods and software tools are employed to model data flows, simulate system behavior, and optimize performance under realistic constraints. Experimental methods, including controlled testing and user feedback, are used to refine the design and ensure it aligns with the client's requirements.

F.1 Computational Methods

Computational methods are central to the evaluation and refinement of the system. Data preprocessing and modeling are handled using Python libraries such as Pandas and NumPy, ensuring the integrity and usability of clickstream data. Machine learning models will be trained using TensorFlow or scikit-learn to generate real-time insights, with a focus on accuracy, scalability, and responsiveness. These models will be tested against pre-defined datasets to validate their performance before integration into the final system.

Interactive data visualizations are an essential component of the project and will be developed using tools like Altair or Plotly. These visualizations are designed to provide users with actionable insights in real time, directly addressing the client's needs for dynamic, user-friendly interfaces. High-fidelity prototypes of the system will also be created using Figma or Adobe XD to simulate workflows and gather user feedback on interface design.

JIRA integration will be validated through API testing using Postman and custom Python scripts to ensure smooth interaction between the system and external tools. This computational approach ensures that the system architecture is robust and capable of handling real-world scenarios.

F.2 Experimental Methods

Experimental methods will complement computational approaches by validating the system under realistic conditions. A controlled testing environment will be set up using Docker containers to replicate production scenarios. These environments will enable the simulation of diverse clickstream behaviors and JIRA interaction workflows. Metrics such as system uptime, response times for query execution, and user satisfaction will be monitored during these tests.

Usability testing will play a critical role in refining the design. Feedback will be gathered during system demonstrations, with observations and surveys used to identify areas for improvement. The usability tests will evaluate the system's interface, navigation, and responsiveness, ensuring it meets the client's expectations.

F.3 Architecture/High-level Design

The system architecture will be designed for modularity and scalability, ensuring that it can be expanded or modified to accommodate future requirements. The front end will be developed using React.js for dynamic and responsive user interfaces, while the back end will utilize Flask or FastAPI to manage requests, data processing, and integration logic. PostgreSQL will serve as the primary database, securely storing clickstream data, system logs, and user configurations. RESTful APIs will be implemented to facilitate communication between the backend, visualization tools, and external integrations such as JIRA.

F.5 Validation Procedure

The final design will be validated through a structured plan to confirm its adherence to client needs and design specifications. A demonstration of the working prototype will be scheduled for early April, during which the system's real-time data visualization, responsive design, and JIRA integration will be showcased. Client feedback will be captured through structured interviews and observation notes, with a focus on usability and functionality.

Feedback will also be solicited from other stakeholders through surveys, gathering quantitative and qualitative insights into the system's performance. A comprehensive final report will document the system's adherence to design objectives, including computational models and experimental results. The report will highlight key performance metrics, such as data accuracy, system reliability, and user satisfaction.

Validation tools, such as Python scripts and visualization software, will confirm data accuracy and system responsiveness. API logs will validate successful interactions with JIRA, while usability testing will ensure the interface meets user expectations. This iterative validation process ensures that the system is robust, user-friendly, and aligned with the client's vision.

Section G. Results and Design Details

G.1 Modeling Details

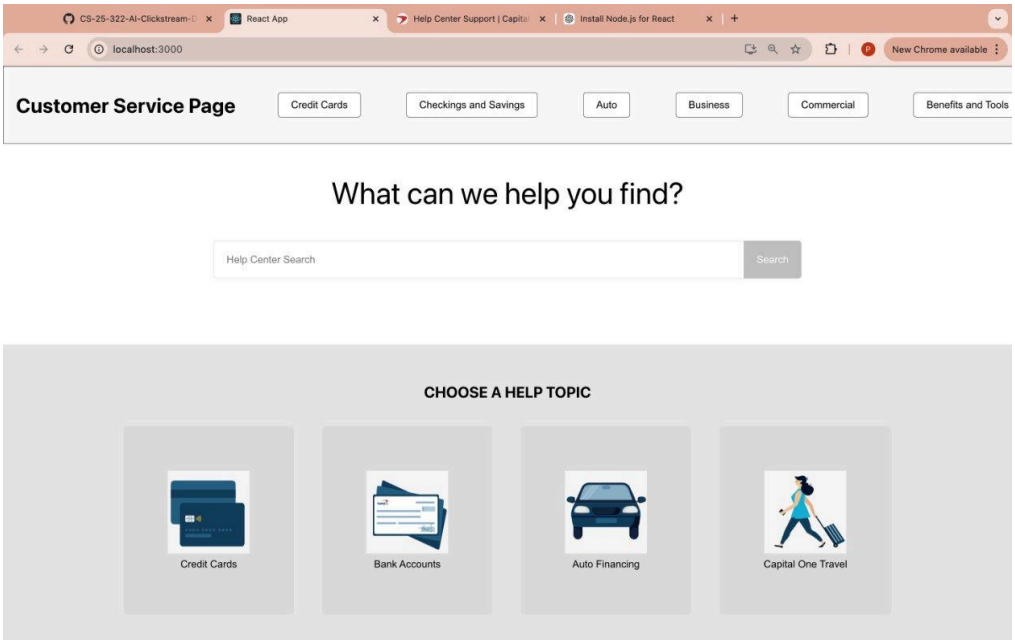
We plan to utilize OpenAI's model for our project due to its advanced ability to understand and process language more effectively than other AI models. For instance, if we provide OpenAI with the words "cat," "dog," and "cap," and ask which pair is most closely related, it would identify "cat" and "dog" as the closest, since both are animals. In contrast, other AI models might incorrectly identify "cat" and "cap" as the closest pair due to the minor spelling similarity. We believe that OpenAI's superior understanding of words will significantly enhance our ability to analyze Clickstream data.

Currently, we do not intend to use a pretrained model, as we aim to train an AI that we built using our own dataset. However, if our efforts to develop a custom model prove to be inadequate, we will explore pretrained alternatives. Based on our research, we see two potential approaches. The first involves utilizing a framework like PyTorch, which would allow us to build a custom model tailored to our data. This approach would leverage an existing foundational model, requiring us to focus on training it to analyze our specific dataset. The second option is to use a platform like Clickstream.io, which is really good at analyzing and visualizing Clickstream data, and then develop a custom AI to process the data from Clickstream.io and generate actionable insights.

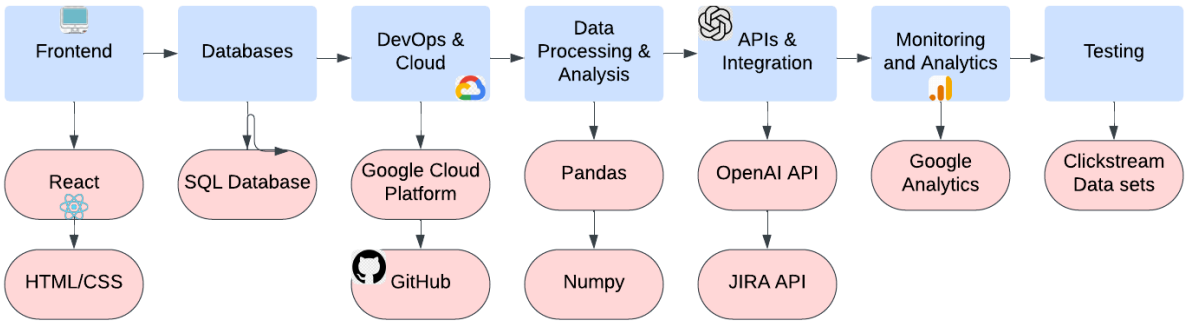
G.2 Experimental Designs

For the frontend, we plan to use React, as it is the framework with which we are most familiar. The UI design of the website will be inspired by the Capital One help page, as we aim to create a seamless user experience that facilitates a smooth transition in the event that Capital One approves our prototype. Additionally, we will integrate Google Analytics into the frontend to assist in analyzing Clickstream data.

Users will be able to interact with the website through various features, such as buttons that redirect to new pages, dropdown menus for selecting options, and text input fields. These interactions will generate Clickstream data, which will be used to train the AI. The data will be stored in an SQL database, in conjunction with Google Cloud Platform, to ensure secure and scalable storage. Our goal is to leverage this data to enhance the website's user-friendliness and overall efficiency.



G.3. Final Design Details/Specifications



Section H. Societal Impacts of Design

H.1 Public Health, Safety, and Welfare

Our project enhances customer service by delivering faster and more accurate responses through AI-driven analysis of clickstream data. By reducing Average Handle Time and improving platform efficiency, we're streamlining user interactions and contributing to a more stress-free customer experience.

To ensure safety and trust, our design adheres to these data privacy standards:

- Data Encryption: All clickstream data is encrypted both during transmission and when stored. This ensures that sensitive information remains secure and inaccessible to unauthorized individuals.
- Privacy Compliance: We follow Capital One's privacy policies, ensuring no Personally Identifiable Information (PII) is stored or exposed. This safeguards users' confidentiality while complying with data protection laws like GDPR and CCPA.

These measures prioritize user safety and build trust, contributing to a secure and reliable platform for all users.

H.2 Societal Impacts

Our project simplifies how users interact with digital banking platforms, making them more accessible and user-friendly, even for individuals who may lack technical expertise. By addressing common challenges and improving platform features, we're creating a smoother experience that meets the needs of a diverse audience. Automation plays a significant role by taking over repetitive tasks, freeing banking teams to focus on creative problem-solving and addressing more complex challenges. However, we acknowledge that the shift to more digital operations will change employee roles. This highlights the growing importance of training programs to help team members acquire new skills and adapt to modern industry demands.

H.3 Environmental Impacts

Our project leverages cloud computing to reduce reliance on physical infrastructure, which helps lower our environmental footprint. By minimizing the need for energy-intensive hardware, we reduce waste and contribute to a more sustainable approach. However, real-time data processing requires significant energy resources. To mitigate this, we're implementing energy-efficient algorithms and choosing cloud providers that prioritize renewable energy sources. By considering sustainability from the design stage, we minimize negative environmental impacts while maintaining the performance and reliability of our system.

H.4 Global Impacts

Our project's impact goes beyond banking by providing a model for organizations worldwide to harness clickstream data for smarter decision-making. By emphasizing data privacy and compliance, it establishes a global benchmark for ethical AI use, creating trust and encouraging responsible innovation. As industries continue to adopt digital systems, this project demonstrates how technology can drive progress while maintaining ethical and privacy standards.

Section I. Cost Analysis

A comprehensive cost analysis was conducted to account for all expenditures related to the development of this project. The analysis includes the costs associated with software tools, hardware components for experimental testing, and labor. If the system is to be scaled as a commercial product, production cost estimates are also considered, including fixed capital, raw materials, and labor costs for manufacturing and packaging.

Current Expenditures
Below is the breakdown of the costs incurred to date for the design and implementation phases:

1. Software Tools
    a. Altair, Plotly, and TensorFlow Libraries: Free (open-source)
    b. Postman (API Testing): Free (Basic Plan)
2. Hardware Components
    a. Development Workstations: Existing systems used, no additional cost
    b. Docker Hosting (Cloud Environment): Estimated $25/month (for 3 months) = $75
3. Miscellaneous
    a. JIRA API License (Developer Access): Included with client's existing account, no additional cost

Total Costs to Date: $75

Bill of Materials for Experimental Set-Up
Since the project relies primarily on software and cloud infrastructure, a detailed Bill of Materials is not extensive. However, for the controlled testing environment, the following costs are associated with cloud resources:

Component	Vendor	Unit Cost	Quantity	Total Cost	Delivery Time	Date Received
Cloud Compute instances	AWS/Google	$25/month	3 months	$75	January 13th	N/A

Production Cost Estimate
Should the system be scaled into a commercial product, the following production cost estimates are considered:

1. **Fixed Capital Costs**
    a. Cloud Infrastructure: $100/month (based on required scaling for 1000 active users)
    b. Server Maintenance: $200/month

2. **Raw Materials**
    a. Additional Software Licenses: $150 (yearly)
3. **Labor Costs**
    a. Software Engineers (2): $40/hour, 10 hours/week, 4 weeks/month = $3,200/month
    b. UI/UX Designers (1): $35/hour, 8 hours/week, 4 weeks/month = $1,120/month
4. **Manufacturing and Packaging**

- As the product is a digital platform, manufacturing and packaging costs are minimal. However, initial marketing materials for deployment and user onboarding may cost an additional $500.

Total Estimated Monthly Costs for Commercial Product

- Fixed Capital Costs: $300
- Raw Materials: $150
- Labor Costs: $4,320
- Marketing and Miscellaneous: $500
- Estimated Monthly Total: $5,270

This preliminary cost analysis will be updated in subsequent reports to include actual costs incurred during final testing and deployment phases, as well as additional production cost estimates based on client requirements and system scaling considerations.

Section J. Conclusions and Recommendations

The design process for this project began with the central question: How can we leverage AI to enhance the efficiency of a webpage? Following this, we identified the necessary resources and established a timeline to guide the project. The first critical step was obtaining data for the AI model to learn from, and we determined that Clickstream data would be an ideal source. Clickstream represents the sequence of links clicked by a user on a website, providing valuable insights into user behavior. However, Clickstream requires a website to gather this data, and since we do not have access to an existing site, we will need to develop our own.

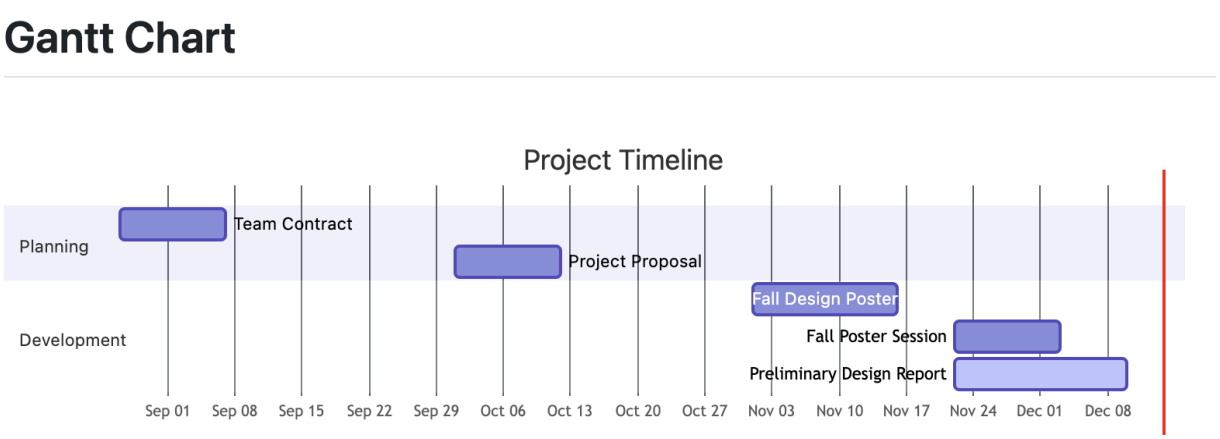
Once the AI is trained using the Clickstream data, the next goal is for the AI to autonomously create new epics and stories within JIRA, a team management tool, which can then be reviewed by teams for potential implementation. Epics and stories are the terms used in JIRA to categorize and track tasks.

This plan represents the core concept behind our project design. However, we encountered challenges when deciding on the optimal resources and technologies for implementation. To address these issues, we met with mentors from both VCU and Capital One, whose advice has been extremely useful in refining our approach. With their advice, we aim to meet the project's primary goals and objectives effectively.

Looking ahead, there is significant potential for further development, particularly in the application of AI. While websites and JIRA integrations may have limitations in prototype form, AI has the capacity for continuous improvement. An interesting question to explore in future iterations is how the AI's suggestions might differ if it were trained using alternative methods. This would require extensive testing to determine the most effective approach. If the AI model we develop does not meet our expectations, the project can be continued by another group. This team would take over our website, either collecting additional data for further AI training or employing a different approach to refine the AI's performance. Our resources, including code and documentation, will be made available through GitHub for easy handover.

The key milestones achieved thus far include the completion of the Team Contract, Project Proposal, and the development of a basic website. For the next semester, our focus will be on integrating Clickstream data into the website and advancing the AI functionality.

Appendix 1: Project Timeline



Appendix 2: Team Contract (i.e. Team Organization)

Step 1: Get to Know One Another. Gather Basic Information.

Task: This initial time together is important to form a strong team dynamic and get to know each other more as people outside of class time. Consider ways to develop positive working relationships with others, while remaining open and personal. Learn each other's strengths and discuss good/bad team experiences. This is also a good opportunity to start to better understand each other's communication and working styles.

Team Member Name	***Strengths each member bring to the group***	***Other Info***	***Contact Info***
Bindi Patel	Outgoing, willing to hear others' ideas, organized, JAVA, Python	interested in cybersecurity, has worked on web design (HTML, CSS, JavaScript)	(540)905-3849 patelb10@vcu.edu
Priya Choudhary	communication, organized, timely, C, Python, Java, PyCharm	Experience with front-end applications, mostly proficient in C and Linux environments, but has experience with Java and python	(703)731-5361 choudharyp2@vcu.edu
Carissa Trieu	Organization, Java, Python, Communication	Interest in data science, experience with Python from VIP, also comfortable with Java	(804)664-5277 trieuc3@vcu.edu
Ivan Emdee	Java, Flexible, Making diagrams,	Hoping to become a software engineer, in the process of learning other languages	(804)245-2593 emdeein@vcu.edu

Other Stakeholders	***Notes***	***Contact Info***
Thomas Gyeera	To get started, I would like for you guys to do some research on articles on related topics and previous research that has been done on this topic. summarize them and understand them so that they can help with your project proposal. you can expect research to also be sent from me to assist. you can expect to see the syllabus by	gyeeratw@vcu.edu

	the end of the weekend. So far I am happy with how prepared you guys were for the meeting today.	
Mahesh Nair Tyler Jordan Emily Croxall	Initial meetings with the team have shown us that you all have a lot of energy and are very engaged with the project. The team came prepared with questions and left the meeting with a clear path forward and action items.	mahesh.bahulleyannair@capitalone.com tyler.jordan@capitalone.com emily.croxall@capitalone.com

Step 2: Team Culture. Clarify the Group's Purpose and Culture Goals.

Task: Discuss how each team member wants to be treated to encourage them to make valuable contributions to the group and how each team member would like to feel recognized for their efforts. Discuss how the team will foster an environment where each team member feels they are accountable for their actions and the way they contribute to the project. These are your Culture Goals (left column). How do the students demonstrate these culture goals? These are your Actions (middle column). Finally, how do students deviate from the team's culture goals? What are ways that other team members can notice when that culture goal is no longer being honored in team dynamics? These are your Warning Signs (right column).

Resources: More information and an example Team Culture can be found in the Biodesign Student Guide "Intentional Teamwork" page (webpage | PDF)

Culture Goals	***Actions***	***Warning Signs***
Being on time	Text in the group chat to remind/decide on meeting times	- text if a date/time doesn't work for you or if you will be missing a meeting. - your responsibility to get updated on what was missed in the meeting
Thursday's at 6 PM will be dedicated to a group status update	- Follow agile principles: update what was done during the week, and what you expect to get done the next week, and identify any roadblocks that you need assistance with. - Set reasonable deadlines and note when an extension is needed	- Student shows up for weekly meeting with no considerable work done - The student doesn't update the group
records for each group member	- a Google doc that will be updated each meeting with information on what each team member plans/has accomplished - every meeting the scribe is rotated	- The scribe forgets to update the doc

Step 3: Time Commitments, Meeting Structure, and Communication

Task: Discuss the anticipated time commitments for the group project. Consider the following questions (don't answer these questions in the box below):

- What are reasonable time commitments for everyone to invest in this project?
- What other activities and commitments do group members have in their lives?
- How will we communicate with each other?
- When will we meet as a team? Where will we meet? How Often?
- Who will run the meetings? Will there be an assigned team leader or scribe? Does that position rotate or will same person take on that role for the duration of the project?

Required: How often you will meet with your faculty advisor advisor, where you will meet, and how the meetings will be conducted. Who arranges these meetings?
See examples below.

Meeting Participants	***Frequency*** ***Dates and Times / Locations***	***Meeting Goals*** ***Responsible Party***
Students Only	As Needed, On Facetime or normal phone call, Zoom if screen sharing	Update group on day-to-day challenges and accomplishments
Students Only	Every Thursday, in the library during normal lab time (if possible)	Actively work on the project
Students + Faculty advisor	Every Friday at noon (Room E4242)	Update the faculty advisor and discuss any questions
Project Sponsor - Tyler Jordan Emily Croxall	Every Friday at noon - can cancel if no topics are needed We will create a recurring Zoom meeting and send an invite to the team.	Help the team with project requirements and possibly technical questions throughout the duration of the project.

Step 4: Determine Individual Roles and Responsibilities

Task: As part of the Capstone Team experience, each member will take on a leadership role, ***in addition to*** contributing to the overall weekly action items for the project. Some common leadership roles for Capstone projects are listed below. Other roles may be assigned with approval of your faculty advisor as deemed fit for the project. For the entirety of the project, you should communicate progress to your advisor specifically with regard to your role.

- **Before meeting with your team**, take some time to ask yourself: what is my "natural" role in this group (strengths)? How can I use this experience to help me grow and develop more?
- **As a group,** discuss the various tasks needed for the project and role preferences. Then assign roles in the table on the next page. Try to create a team dynamic that is fair and equitable, while promoting the strengths of each member.

Communication Leaders

Suggested: Assign a team member to be the primary contact for the client/sponsor. This person will schedule meetings, send updates, and ensure deliverables are met.

Suggested: Assign a team member to be the primary contact for faculty advisor. This person will schedule meetings, send updates, and ensure deliverables are met.

Common Leadership Roles for Capstone

1. **Project Manager:** Manages all tasks; develops overall schedule for project; writes agendas and runs meetings; reviews and monitors individual action items; creates an environment where team members are respected, take risks and feel safe expressing their ideas.
   Required: On Edusourced, under the Team tab, make sure that this student is assigned the Project Manager role. This is required so that Capstone program staff can easily identify a single contact person, especially for items like Purchasing and Receiving project supplies.
2. **Logistics Manager:** coordinates all internal and external interactions; lead in establishing contact within and outside of organization, following up on communication of commitments, obtaining information for the team; documents meeting minutes; manages facility and resource usage.
3. **Financial Manager:** researches/benchmarks technical purchases and acquisitions; conducts pricing analysis and budget justifications on proposed purchases; carries out team purchase requests; monitors team budget.
4. **Systems Engineer:** analyzes Client initial design specification and leads establishment of product specifications; monitors, coordinates and manages integration of sub-systems in the prototype; develops and recommends system architecture and manages product interfaces.
5. **Test Engineer:** oversees experimental design, test plan, procedures and data analysis; acquires data acquisition equipment and any necessary software; establishes test protocols and schedules; oversees statistical analysis of results; leads presentation of experimental finding and resulting recommendations.
6. **Manufacturing Engineer:** coordinates all fabrication required to meet final prototype requirements; oversees that all engineering drawings meet the requirements of machine shop or vendor; reviews designs to ensure design for manufacturing; determines realistic timing for fabrication and quality; develops schedule for all manufacturing.

Team Member	*Role(s)*	*Responsibilities*
Priya Choudhary	Project Manager	• Keep a detailed record of meeting notes and share them with the group • creates an environment where team members are respected • develops the overall schedule for the project • writes agendas and runs meetings
Carissa Trieu	Test Engineer	• Oversee project design, test plan, procedures, and data analysis • Establish test protocols and schedules • Lead presentation of recommendations and results
Bindi Patel	Logistics & Financial Manager	• Primary contact with advisors (updating them, etc) • Record meeting minutes • Handles the research/comparison of purchases • Responsible for keeping track of the team budget
Ivan Emdee	Systems Engineer	• Analyze the client's initial designs to come up with a prototype • Oversee the development of subsystems • Recommend system architecture and manage product interface

Step 5: Agree to the above team contract

Team Member: Priya Choudhary *Signature: Priya Choudhary*

Team Member: Carissa Trieu *Signature: Carissa Trieu*

Team Member: Bindi Patel *Signature: Bindi Patel*

Team Member: Ivan Emdee *Signature: Ivan Emdee*

References

[1] VCU Writing Center. (2021, September 8). *APA Citation: A guide to formatting in APA style.* Retrieved September 2, 2024. https://writing.vcu.edu/student-resources/apa-citations/

[2] Teach Engineering. *Engineering Design Process*. TeachEngineering.org. Retrieved September 2, 2024. https://www.teachengineering.org/populartopics/designprocess

[3]J. Electrical Systems 20-2 (2024):1600-1608.Real-Time Clickstream Analytics with Apache. Retrieved October 10, 2024. https://www.researchgate.net/profile/Tulshihar-Patil/publication/382193441_Real-Time_Clickstream_Analytics_with_Apache/links/6691464a3e0edb1e0fe0ce1c/Real-Time-Clickstream-Analytics-with-Apache.pdf

[4]Hanamanthrao, R., & Thejaswini, S. (2017). Real-time clickstream data analytics and visualization. 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). https://doi.org/10.1109/rteict.2017.8256978

[5]Ehsan, A., Abuhaliqa, M. A. M. E., Catal, C., & Mishra, D. (2022). RESTful API Testing Methodologies: Rationale, Challenges, and Solution Directions. Applied Sciences, 12(9), 4369. https://doi.org/10.3390/app12094369

[6]Yusof, M. K., Man, M., & Ismail, A. (2022). Design and Implement of REST API for Data Integration. 2022 International Conference on Engineering and Emerging Technologies (ICEET). https://doi.org/10.1109/iceet56468.2022.10007414

[7]M. Gumber, A. Jain and A. L. Amutha, "Predicting Customer Behavior by Analyzing Clickstream Data," *2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICCCSP52374.2021.9465526.

[8]Liu Y, Fan S, Xu S, Sajjanhar A, Yeom S, Wei Y. Predicting Student Performance Using Clickstream Data and Machine Learning. Education Sciences. 2023; 13(1):17. https://doi.org/10.3390/educsci13010017

[9]Ulitzsch, E., Ulitzsch, V., He, Q. et al. A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks. Behav Res 55, 1392–1412 (2023). https://doi.org/10.3758/s13428-022-01844-1

[10]Severeyn, E., Alexandra La Cruz, Matute, R., & Estrada, J. (2023). Neural Networks for Customer Classification Through Clickstream Analysis. https://doi.org/10.1109/etcm58927.2023.10309081