

# **Asklytics**

***An LLM for Website Analytics with Sentiment Analysis of Google  
Analytics 4 (GA-4) & Google Maps Reviews Data***

**Course: COMP 4495**

**Applied Research Project**

**Section: 001**

**Team Lead**

Tamoor Haider Aslam (300367290)

**Team Members**

Mohamed Nuskhan Mohamed Niyas (300368621)

**Date of Submission:**

26.01.2025

## Table of Contents

<b>INTRODUCTION .....</b>	<b>1</b>
<i>Domain .....</i>	<i>1</i>
<i>Background &amp; Context.....</i>	<i>1</i>
<i>Problems to Solve .....</i>	<i>1</i>
<i>Existing Tools .....</i>	<i>1</i>
<i>Hypothesis, Assumptions &amp; Benefits .....</i>	<i>3</i>
<b>PROPOSED RESEARCH PROJECT .....</b>	<b>3</b>
<i>Research Design and Objectives:.....</i>	<i>3</i>
<i>Methodology and Justification:.....</i>	<i>4</i>
<i>Data Collection:.....</i>	<i>4</i>
<i>Technologies:.....</i>	<i>5</i>
<i>Expected Results: .....</i>	<i>5</i>
<b>RIIPEN EXTERNAL PARTNERS OR AFFILIATES .....</b>	<b>5</b>
<b>PROJECT PLANNING AND TIMELINE .....</b>	<b>5</b>
<b>PROJECT CONTRACT .....</b>	<b>8</b>
<b>WORK DATE/HOURS LOGS .....</b>	<b>8</b>
<b>CLOSING AND REFERENCES.....</b>	<b>9</b>

# Introduction

## Domain

In the modern era of data-driven decision-making, organizations increasingly rely on web analytics to understand user behavior, optimize website performance, and enhance customer satisfaction. Google Analytics 4 (GA4) serves as a powerful tool for gathering web traffic data, offering insights into metrics such as bounce rates, user sessions, and engagement patterns. However, despite its utility, interpreting this data can be overwhelming for non-technical users, limiting its accessibility and practical application.

## Background & Context

This project addresses the challenge of making website analytics user-friendly by developing an AI-powered chatbot capable of answering natural language queries. Beyond the standard analytical insights, the chatbot will also integrate sentiment analysis capabilities to classify customer feedback into positive, neutral, or negative sentiments. This dual approach not only simplifies data interpretation but also provides actionable insights for businesses to improve their offerings and customer engagement strategies.

## Problems to Solve

The key problems this project aims to solve include:

1. The steep learning curve associated with understanding and using GA4 data.
2. The absence of conversational interfaces in current analytics tools.
3. The lack of integrated sentiment analysis in interpreting user feedback along with the real time impact on user interest in business's website.

## Existing Tools

Existing research and tools focus heavily on data visualization and automated reporting but rarely incorporate natural language interfaces tailored to specific analytics questions. By leveraging Natural Language Processing (NLP) models, sentiment analysis techniques, and

predictive analytics, this project fills a critical gap in website performance analysis and feedback interpretation.

## Literature Review

1. This paper highlights how integrating AI and data analytics is revolutionizing competitive intelligence, enabling businesses to gather, analyze, and act on market insights more effectively. AI-driven technologies like machine learning, natural language processing, and predictive analytics allow organizations to uncover patterns, predict trends, and automate tasks like competitor monitoring and sentiment analysis. Real-time analytics and intuitive visualization tools further enhance decision-making and resource allocation, helping businesses adapt dynamically to changing market conditions. While the adoption of these technologies presents challenges such as data privacy, integration issues, and skill gaps, the paper provides strategic recommendations to address them. Case studies from various industries demonstrate the benefits of AI-powered competitive intelligence, including driving innovation, improving customer engagement, and optimizing strategies

*Jack, Harper. (2024). Leveraging AI and Data Analytics: Revolutionizing Competitive Intelligence for Market Insights. 10.13140/RG.2.2.25148.76166.*

2. The rise of big data and advanced analytics has transformed decision-making, enabling organizations to shift from reactive to predictive strategies. Predictive analytics uses machine learning, AI, and statistical models to analyze historical and real-time data, uncover patterns, forecast trends, and optimize decisions. It is widely applied in industries like healthcare, finance, manufacturing, and retail, enhancing outcomes such as improved patient care, fraud prevention, and predictive maintenance. However, challenges such as data quality, integration complexities, and ethical concerns persist. This paper explores the transformative role, applications, challenges, and prospects of predictive analytics, emphasizing its

importance in fostering resilience, sustainability, and competitive advantage in a data-driven world.

*Nyoni, Rumbidzai. (2025). Harnessing Data Analytics for Predictive Insights: Advancing Decision- Making with Big Data Innovations. International Journal of Research Publication and Reviews. 6. 2915-2936.*

3. Financial institutions are increasingly using data lakes to manage vast and diverse datasets, enabling advanced fraud detection and prevention. These repositories store structured, semi-structured, and unstructured data, facilitating insights into large-scale financial information. By integrating advanced analytics techniques such as machine learning, anomaly detection, predictive modeling, and natural language processing (NLP), data lakes empower institutions to identify patterns, predict fraudulent behavior in real time, and take proactive measures. This paper examines the architecture and technologies behind financial data lakes, their practical applications in fraud prevention, and the challenges of implementation. It also provides a roadmap for successful adoption, emphasizing their role in safeguarding assets and maintaining customer trust.

*Arena, Frank & Paul, Justin. (2024). INNOVATIVE ANALYTICS TECHNIQUES FOR FINANCIAL DATA LAKES.*

## Hypothesis, Assumptions & Benefits

Our hypothesis is that an intuitive, AI-driven chatbot combined with sentiment analysis will empower users to derive meaningful insights without requiring technical expertise. This project has the potential to enhance decision-making, reduce the time spent on manual analysis, and improve customer-centric strategies for organizations.

## Proposed Research Project

### Research Design and Objectives:

The overarching goal of this project is to create an interactive chatbot that simplifies website analytics and enhances customer feedback analysis. The specific objectives include:

1. Implementing a natural language interface to process and respond to queries about website performance metrics.
2. Visualizing GA4 data in an accessible and interactive format.
3. Classifying user feedback into sentiment categories (positive, neutral, and negative).
4. Providing predictive insights for future website performance and detecting anomalies in traffic patterns.

### Methodology and Justification:

- *Natural Language Understanding (NLU)*: The chatbot will be powered by a fine-tuned transformer-based model such as BERT, tailored to analytics-related queries. BERT's ability to understand context and semantics ensures accurate responses to user questions.
- *Sentiment Analysis*: User feedback will be analyzed using pre-trained sentiment analysis models such as VADER or `en_core_web_lg`. These models have proven effective in classifying text sentiments, particularly for reviews and social media data.
- *Predictive Analytics*: To forecast website traffic trends, models like ARIMA and Prophet will be used. Anomaly detection will be implemented with Isolation Forest or DBSCAN to identify unusual patterns in the data.
- *Visualization*: Data visualization will be achieved using Chart.js and Plotly to ensure clarity and interactivity, enabling users to explore metrics like traffic sources, engagement rates, and bounce rates.

### Data Collection:

1. *GA4 Data*: Metrics such as user sessions, bounce rates, and page views will be retrieved using the Google Analytics Data API.
2. *Google BigQuery*: Google Merchandise Ecommerce Store publicly available data will be used as training dataset.
3. *User Feedback*: Publicly available datasets containing user reviews will be preprocessed and used for sentiment analysis.

## Technologies:

- *Operating System/Platform:* Linux (Ubuntu 24.04 LTS).
- *Programming Languages:* Python for backend and machine learning tasks, JavaScript for frontend development.
- *Frameworks:* Django (backend) and React.js (frontend).
- *Database:* PostgreSQL for managing user sessions, interaction logs, and derived insights.

## Expected Results:

The expected outcomes of this project are:

1. A conversational AI-powered chatbot that responds accurately to user queries about website analytics.
2. Interactive visualizations that make data exploration intuitive and user-friendly.
3. Sentiment classification of user reviews, identifying patterns and trends in feedback.
4. Predictive models that provide insights into future website performance and flag anomalies in real-time.

## Riipen External Partners or Affiliates

This project does not involve any Riipen external partners or affiliates.

## Project Planning and Timeline

### Timeline and Milestones:

Phase	Milestone	Deliverables	Deadline
1	Topic Research, Projects & Repo Setup	Proposal Submission	Jan 26, 2025

Phase	Milestone	Deliverables	Deadline
2	<b>Analytics:</b> <ul style="list-style-type: none"> <li>NLP Pipeline Implementation,</li> </ul> <b>Development:</b> <ul style="list-style-type: none"> <li>Chat Interface Complete (React)</li> <li>User Mood Meter Complete (React)</li> <li>REST API design (Partial)</li> </ul>	Fine-tuned BERT model for natural language queries	Feb 14, 2025
3	<b>Analytics:</b> <ul style="list-style-type: none"> <li>Sentiment Analysis</li> </ul> <b>Development:</b> <ul style="list-style-type: none"> <li>OAuth2 flow (Complete)</li> <li>Implement RESTAPI endpoints with mock data except OAuth2 (complete)</li> </ul>	Sentiment classification module for user feedback	Feb 28, 2025
4	<b>Analytics:</b> <ul style="list-style-type: none"> <li>Data Visualization</li> </ul> <b>Development:</b> <ul style="list-style-type: none"> <li>Integrating RESTAPI with Frontend (Complete).</li> <li>Adding Plots/Charts generation in React (Partial)</li> </ul>	Integrated Chart.js and Plotly components	Mar 14, 2025
5	<b>Analytics:</b> <ul style="list-style-type: none"> <li>Predictive Analytics</li> </ul> <b>Development:</b> <ul style="list-style-type: none"> <li>Adding Plots/Charts generation in React (Complete)</li> </ul>	Traffic forecasting and anomaly detection models	Mar 28, 2025

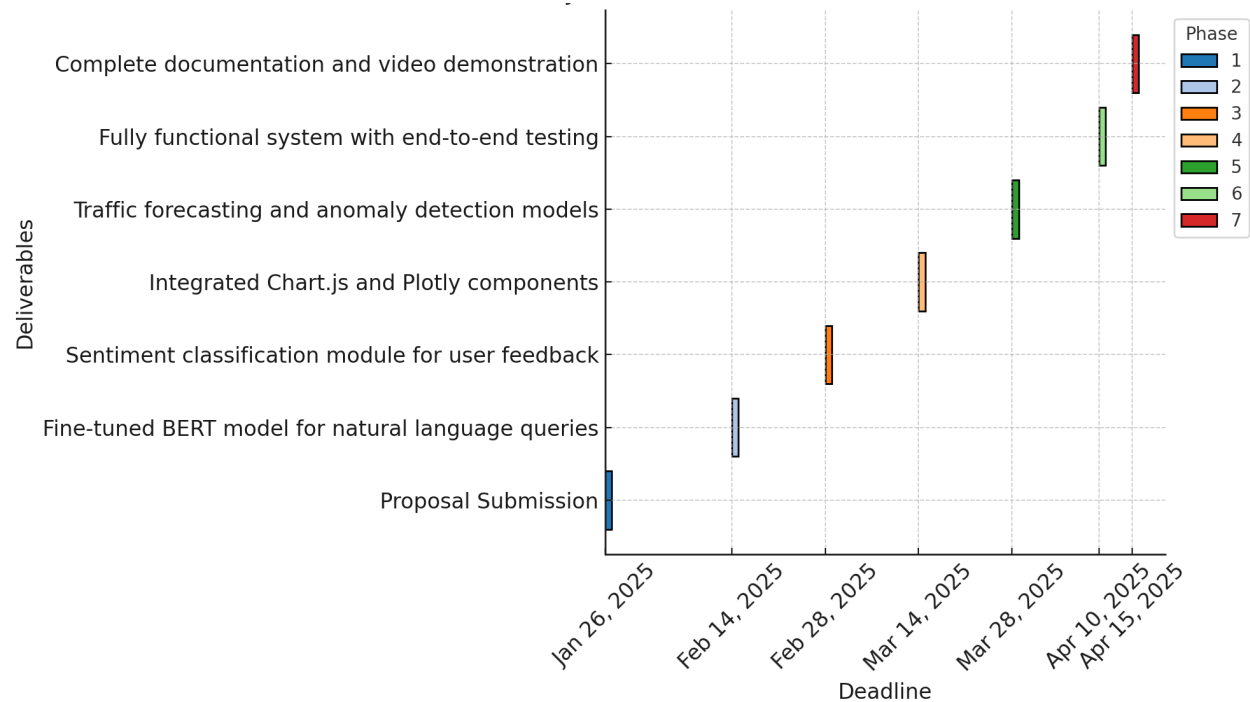


Phase	Milestone	Deliverables	Deadline
6	Integration & Testing	Fully functional system with end-to-end testing	Apr 10, 2025
7	Final Submission	Complete documentation and video demonstration	Apr 15, 2025

### Team Responsibilities:

- **Mohamed Nuskhan Niyas:** Responsible for leading the LLM training, testing, improving & creating NLP pipeline implementation.
- **Tamoor Haider Aslam:** As a Teamlead, Tamoor will be responsible for project management, communication, giving direction, removing blockers etc. Also he will take part in LLM testing, frontend and backend architecture design and development.

### Project Management Chart:



## Project Contract

We, the undersigned, accept & agree with the scope, milestones, and timelines outlined in this proposal. Regular meetings and progress updates will be conducted to ensure the project is completed successfully.

### Digital Signatures:

**Mohamed Nuskhan Niyas:** Mohamed Niyas Mohamed Nuskhan

**Tamoor Haider Aslam:** Tamoor Haider Aslam

## Work Date/Hours Logs

Student Name	Date	Hours	Description of Work Done
Tamoor Aslam	Jan 18, 2025	3	Research on Project Ideas <ul style="list-style-type: none"><li>Audio to sign language avatar video</li></ul>
Mohamed Nuskhan	Jan 20, 2025	3	Researched GA4 API documentation and integration methods.
Tamoor Aslam	Jan 25, 2025	2	Set up repos, worked on proposal doc.
Tamoor Aslam	Jan 26, 2025	2	Formatting, proofreading & referencing. Setup React & Django projects
Mohamed Nuskhan	Jan 22, 2025	1.5	Implemented initial prototype for natural language queries.

## Closing and References

### Acknowledgments:

Special thanks to Douglas College for providing access to resources, tools, and guidance essential for this project.

### References:

1. Devlin, J., et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.
2. Google Developers. (2025). *Google Analytics Data API Documentation*. Retrieved from <https://developers.google.com/analytics/devguides/reporting/data/v1>.
3. Hutto, C., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*.
4. Arena, F., & Paul, J. (2024). Innovative analytics techniques for financial data lakes.
5. Nyoni, R. (2025). Harnessing data analytics for predictive insights: Advancing decision-making with big data innovations. *International Journal of Research Publication and Reviews*, 6, 2915–2936.
6. Jack, H. (2024). Leveraging AI and data analytics: Revolutionizing competitive intelligence for market insights. <https://doi.org/10.13140/RG.2.2.25148.76166>