International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 359-362 www.ijaem.net ISSN: 2395-5252

Data Enrichment Process through User Profiling

Anusha S V, Bhoomika A, Akshatha G S, Bhavana P

Department of Computer Science and Engineering PES Institute of Technology and Management Shivamogga, Karnataka, India

Department of Computer Science and EngineeringPES Institute of Technology and Management Shivamogga, Karnataka, India

Department of Computer Science and Engineering PES Institute of Technology and Management Shivamogga, Karnataka, India

Department of Computer Science and EngineeringPES Institute of Technology and Management Shivamogga, Karnataka, India

Date of Submission: 05-05-2024 Date of Acceptance: 15-05-2024

ABSTRACT—In today's technology driven world user profiles are the virtual representation of each user and they include a variety of user information such as personal, interest and preference data.The main aim of project is to enhance the depth and accuracy to customer information through the strategic integration of external data sources. This project addresses the challenge of enriching the existing customer datasets by identifying and filling critical data gaps. The successful implementation of this project ultimately facilitates to increased customer satisfaction and optimized business strategies.

Keywords—Cloud Computing,User Profiling,DataEnrichment,Data Classification.

I. INTRODUCTION

Data Enrichment Process explains semantic data enrichment that improves the usability. Companies are collecting vast amount of data to gain insight from it to make better decisions and enhance customer experience. If the quality of data collected is poor, no matter how powerful the data analytics is, businesses cannot gain the expected benefits from it. To achieve maximum benefits, data needs to cleaned, organized, and refined with additional information or intelligence. data enrichment process is a strategic initiative aimed at enhancing the quality and utility of raw datasets. It involves augmenting existing data with additional information from diverse sources, refining and expanding its content. By incorporating external data, such as demographic details, geospatial information, or market trends, the enriched dataset becomes more comprehensive and insightful. This refined data is crucial for making informed decisions, conducting thorough analyses, and gaining deeper insights into various aspects of business, research, or any domain relying on accurate and up-to-date information. The data enrichment process is instrumental in unlocking the full potential of datasets, empowering organizations to derive greater value from their information resources. As the world is changing constantly, static data cannot be trusted completely since these do not capture the most recent facts based on which dynamic decisions could be made.

So, to make the right decision based on the most updated information, static data should be enriched with external data. The scope and the requirement of this project is need of the hour as individual's and Company's background verification is becoming very hard to do manually. Since, now a days everything is digitized, our scope is to go out in the digital world and search, filter, and synthesize the information and keep it ready for consumption on demand basis.Enhance user profiles by supplementing existing data with additional information. This could include demographic data (age, gender, location), psychographic data (interests, values, attitudes), behavioral data (purchase history, browsing behavior), or external data sources (third-party data, social media profiles).data enrichment through user profiling enables organizations to better understand their user base, improve targeting and personalization efforts, and ultimately drive business growth and customer satisfaction.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 359-362 www.ijaem.net ISSN: 2395-5252

II. RELATED WORK

Data enrichment through user profiling involves analyzing, retrieving and summarizing the conceptual data to create comprehensive profiles. The data is provided by users directly or data is collected from websites, the fetching method involves acquiring additional data from various sources, initially identifying relevant ones like demographic databases. Summarizing an article involves converting complex information into a more understandable format, preserving its essence and key insights, facilitating comprehension, retention, and dissemination of the main ideas for further profile creation. The text summarizes a conceptual data analysis using ML algorithms and NLP, highlighting key points of the article through Name Entity Recognition (NER) method and providing a user profile through this method. The proposed algorithm aimed to generate user profiles that were representing user's multiple interests. Successful data enrichment with user profiles can enhance user experience, drive business growth, and yield valuable insights by obtaining consent, implementing security measures, and addressing biases. In additional the user profiles are stored securely on a cloud platform called File stack, optimized for scalability and security, ensuring data accessibility and protection. This cloud-based approach facilitates efficient storage and retrieval, enabling personalized experiences at scale.

III. PROPOSEDSYSTEM

The process of enriching data involves data collection, preprocessing, text summarization, data extraction, data integration, validation, analysis, and data maintenance. Data collection entails gathering relevant articles and audio files either manually or through automated means like web scraping or API integration. Preprocessing ensures consistency and accuracy by cleaning and formatting the collected data, removing irrelevant information, and correcting errors. Text summarization condenses the content of articles or audio transcripts into concise summaries through extractive or abstractive methods using NLP techniques. Data extraction involves extracting details like entities, keywords, and sentiments from the summarized text using NLP techniques such as named entity recognition and sentiment analysis. Data integration merges the extracted data with existing datasets, cleaning, deduplicating, and resolving inconsistencies. Validation compares new data with external sources to ensure accuracy. Analysis utilizes data visualization, statistical analysis, and machine learning to derive insights and inform decisions. Finally, data maintenance involves regular updates, monitoring external sources, updating extraction processes, and performing validation checks to maintain relevance and accuracy over time. These steps collectively ensure comprehensive and reliable enriched data for decision-making and insights.

IV. METHODOLOGY

This represents the contemporary methodology for enriching data through user profiling

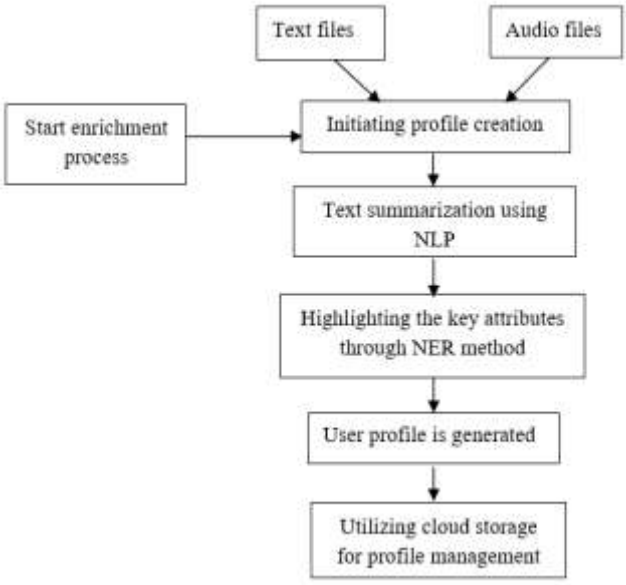

Fig1:SystemArchitecture

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 359-362 www.ijaem.net ISSN: 2395-5252

The process starts with user-provided data or web scraping, leveraging various methods to gather additional data from diverse sources, initially targeting relevant ones like demographic databases. Summarizing articles involves converting complex information into an easily digestible format, preserving key insights for better understanding and dissemination, aiding in profile creation.ML algorithms and NLP techniques analyze text, using NER to highlight key points and create user profiles representing multiple interests.Successful enrichment enhances user experience, drives business growth, and yields insights, ensuring consent, security, and bias mitigation. User profiles are securely stored on the cloud platform File stack, optimized for scalability and security, facilitating efficient storage, retrieval, and personalized experiences.

V. RESULT

The results of the project demonstrate the efficacy of utilizing conceptual data from articles to enhance user profiling

Fig2:Profile creation

The project utilized NLP techniques and NER method, for extract conceptual data from articles and enriching user profiles with demographic details, interests, and behavior patterns inferred to create detailed user profiles. The profiles provided insights into users' preferences, enabling personalized experiences and targeted content delivery. This approach demonstrates the effectiveness of leveraging conceptual data for user profiling, with implications for applications like targeted marketing and personalized recommendations

CONCLUSION

The successful execution of the Data enrichment project not only address immediate data challenges but also sets the stages for a more agile, customer-centric and strategically aligned organization. The enriched data serves as a valuable asset, propelling the organization towards sustained success and excellence in customer satisfaction.

FUTURE WORK

The work could be expanded to include a user profile system, Where can be used for marketing purpose and create the profile according to the person's interest.

REFERENCES

[1] Salahuddin A Azad, Saleh Wasimi, ABM Shawkat Ali Research to Business Data Enrichment: Issues and Challenges, 2018.
[2] M. Toussant, The Importance of Human-Curated Data Enrichment of Big Data Analysis [Blog], Mar. 2018.
[3] P. Wang, W. Hey, R. Shea, J. Wang, and E. Wu, "Deeper: A Data Enrichment System Powered by Deep Web," in Proc. of the 2018.
[4] H. Isah, T. Abughofa, S. Mahfuz, D. Ajerla, F. H. Zulkernine, and S. Khan, "A survey of distributed data stream processing frameworks, 2019.
[5] Fhafa, B. Kilic, and P. Krause, ''Evaluation of IoT stream processing at edge computing layer for semantic data enrichment,'' Apr.

2020,
[6] Fatma Özcan, Chuan Lei, Abdul Quamar, Vasilis Efthymio, Data Management for End-To-End Machine Learning, 2021.
[7] Dominik Scheinert, Fabian Casares, Morgan K Geldenhuys, Kevin Styp-Rekowski, Odej Kao 2023