International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AutoStream: Your All-in-One Data Science Assistant Using LLM

Sowmiya R
Assistant Professor, Department of Artificial Intelligence and Data Science SNS College of Engineering
Coimbatore, India
sowmiyar272@gmail.com

P. Sumathi
Head of Department, Department of Artificial Intelligence and Data Science SNS College of Engineering
Coimbatore, India
psumathi.it@gmail.com

Lavanya A K
Student, Department of Artificial Intelligence and Data Science SNS College of Engineering
Coimbatore, India
lavanya.ak..ad.2021@snsce.ac.in

Mohammed Afsal V P *Student, Department of Artificial Intelligence and Data Science SNS College of Engineering*
Coimbatore, India
mohdafzaal6097@gmail.com

Mohammed Rahees P *Student, Department of Artificial Intelligence and Data Science SNS College of Engineering*
Coimbatore, India
mohammedrahees4850@gmail.com

Gokul C H
Student, Department of Artificial Intelligence and Data Science SNS College Of Engineering
Coimbatore, India
gokulkannang47@gmail.com

ABSTRACT: ***In today's data-driven world, businesses and analysts face significant challenges in extracting actionable insights from large datasets, particularly CSV files, without advanced technical expertise. AutoStream addresses this gap by offering an intuitive, local-first platform designed to empower users of all skill levels. By simplifying data exploration, the platform allows seamless uploading of CSV files, automatic generation of key insights, and interactive visualization tools for in-depth analysis — all without writing a single line of code.Built with user-centric design and extensibility in mind, AutoStream eliminates the complexities of manual data processing and empowers faster, smarter decision-making. Features like natural language data querying (powered by Mistral + PandasAI), one-click EDA reports (via Sweetviz), AutoML capabilities (with PyCaret), and SHAP-based explainability tools help users navigate their data with ease and confidence.With everything running securely on the user's local machine, Auto Stream is ideal for sensitive or enterprise environments where privacy is paramount. This innovative platform not only saves time but also bridges the gap between raw data and strategic insight.***

Would you like further refinements or to expand it?

Keywords: Solution Overview, Core Feature,Problem Addressed, User- centric diesgn

1. INTRODUCTION

In today's fast-paced business environment, organizations face significant challenges in transforming large datasets into actionable insights, particularly without strong technical expertise. The complexity of manually analyzing CSV files, extracting valuable patterns, and generating reports can lead to inefficiencies and missed opportunities. As businesses strive for better decision-making and operational efficiency, the need for intuitive, user-friendly data analysis tools has never been greater. **AutoStream** is designed to address these challenges by offering a streamlined, data-driven platform that simplifies the process of data exploration and insight generation. By allowing users to upload CSV files effortlessly, AutoStream automatically processes and analyzes the data, providing real-time insights through interactive visualizations and reports. This solution eliminates the need for manual data manipulation and empowers teams, both technical and non-technical, to make informed decisions with confidence. This paper explores the design, features, and benefits of AutoStream, which aims to revolutionize how businesses interact with data. By bridging the gap between data complexity and actionable intelligence, AutoStream fosters a culture of data-driven decision-making, helping organizations optimize operations, improve productivity, and stay competitive in an increasingly data-centric world.

II. RELATED SYSTEM

AutoStream addresses key limitations commonly found in existing data analysis systems, which often hinder effectiveness and ease of use. Traditional platforms typically rely on manual data entry, increasing the risk of errors and inconsistencies that compromise data quality. This dependency not only affects the accuracy of insights but also delays the process of transforming raw data into actionable information.

By contrast, AutoStream automates data handling from the moment a CSV file is uploaded, significantly reducing manual input and the potential for human error. Furthermore, many conventional systems lack real-time analysis capabilities, preventing users from accessing up-to-date insights crucial for timely decision-making. AutoStream overcomes this with its built-in, real-time interactive analysis features, allowing users to `ask questions in plain English and receive immediate visual or statistical feedback.

Fig 1. An existing OEE system with its featured dashbaord

Another significant limitation of traditional data analysis systems is their inability to seamlessly integrate with other enterprise platforms such as Customer Relationship Management (CRM) tools. Without this integration, these systems operate in isolation, creating data silos that hinder a comprehensive understanding of business operations. This lack of connectivity prevents businesses from performing cross-functional analysis, leading to suboptimal decision-making and missed opportunities for optimization. Additionally, traditional systems often lack advanced analytics capabilities, limiting their ability to identify trends, detect anomalies, or forecast future needs.AutoStream AutoStream overcomes these challenges by leveraging advanced predictive algorithms that analyze uploaded datasets to identify patterns, predict trends, and provide actionable insights. The system's flexible, open-source architecture enables easy integration with CRM and other enterprise platforms, ensuring a unified view of organizational data and enabling data-driven decisions across all departments. Through real-time natural language querying, automated data profiling, and AutoML-powered forecasting, AutoStream delivers a comprehensive and interactive view

III. PROPOSED SYSTEM

The proposed OEE (Overall Equipment Effectiveness) system introduces a modern, technology-driven approach to equipment monitoring and optimization, addressing the constraints and inefficiencies found in traditional OEE tracking methods. This system integrates IoT (Internet of Things) technology, automation, machine learning, and advanced analytics to provide a continuous, comprehensive view of equipment performance. Unlike legacy systems, which rely heavily on manual data input and lack immediacy, this OEE system's automated data collection and real-time capabilities enable manufacturers to monitor equipment conditions closely and make prompt, informed decisions.A key feature of the system is its real-time monitoring capability, made possible through IoT-enabled sensors installed on equipment. These sensors constantly capture and transmit data relevant to OEE's three core metrics—Availability, Performance, and Quality—providing an accurate, up-to-date view of equipment status. By eliminating delays in data acquisition, the system allows operators and managers to immediately identify and address emerging issues, reducing unplanned downtimes and improving overall productivity.Automation in data collection is another central aspect, as it minimizes the need for manual data entry, which is prone to human error and can compromise data accuracy. Automated data acquisition ensures that the data being analyzed is reliable and consistent, allowing for precise performance evaluation and decision-making. With this system, organizations can access a robust data stream that enhances data integrity, leading to a better understanding of equipment performance and operational patterns.

Fig 2. The dashboard of OEE Working which is shown to employees

Finally, the system incorporates a user-friendly, role-based interface to make data accessible and actionable for departments within the organization.

For non-technical users, such as business stakeholders, **AutoStream** provides user-friendly interfaces that display actionable insights and visualizations, making complex data easily accessible. More advanced users can dive deeper into datasets, exploring granular details and running sophisticated analyses with the platform's powerful tools.drive operational efficiency.

IV. METHODOLOGY

SOFTWARE REQUIREMENTS: **AutoStream** employs a multi-layered architecture designed to efficiently handle large datasets, with a primary focus on ease of use and accessibility for non-technical users. The methodology follows a modular approach, integrating cutting-edge technologies to ensure seamless data extraction, processing, and visualization.

Data Collection (Cameras & Sensors): Cameras and sensors play a critical role in monitoring operational metrics, such as system uptime, downtime, and product quality. Real-time data collected from these sources is automatically integrated into **AutoStream**, allowing for continuous performance monitoring and insights. These data streams are vital for enabling real-time decision-making and operational adjustments, all processed and visualized within the platform.

Data Storage (MongoDB): AutoStream leverages **MongoDB** for optimized data storage, both for structured and unstructured data. MongoDB's NoSQL design allows for rapid retrieval and processing, ensuring that large datasets can be managed efficiently without the constraints of traditional relational databases.

. The system's user-friendly interface and intuitive design allow non-technical users to interact with complex datasets, making it an invaluable tool for improving business outcomes.

Module 1:

AutoStream focuses on **data acquisition** through real-time monitoring via integrated cameras. These cameras are strategically placed to capture critical metrics such as **uptime**, **downtime**, and **product quality**. The data collected is transmitted securely to the backend using real-time communication protocols like **WebSocket** and **MQTT**, ensuring efficient data flow. Once received, the data undergoes **preprocessing** to clean and structure it for further analysis. This step includes noise reduction, data validation, and formatting to ensure accuracy and usability.

The system also performs initial **data analysis** by identifying issues such as downtime or quality defects, and raises alerts in real-time. These alerts are communicated to users via push notifications or emails, and the status is reflected on the **dashboard**. This module ensures that high-quality, real-time data is available for the next stages of the system, driving accurate insights and decisions.

Fig 3. Fire Frame

Module 2:

AutoStream focuses on data processing, where the raw data collected from the cameras is analyzed, processed, and converted into meaningful insights. The backend, built with Node.js and Express.js, plays a central role in this module by handling data processing tasks, calculating Overall Equipment Effectiveness (OEE) scores, and integrating with other systems via APIs.

Once the data is received from the data acquisition layer, it is processed to calculate key performance indicators (KPIs) such as uptime, downtime, and product quality. The backend performs these calculations in real-time, ensuring that the system provides up-to-date insights for decision-making. OEE scores are calculated based on the collected data, which helps businesses monitor equipment efficiency and optimize production processes.

Fig 4. The Login Page of the Application to start the software

Module 3:

AutoStream focuses on the **frontend**, which is responsible for presenting the processed data to users in a visually appealing and interactive manner. Built using **React.js**, the frontend enables users to easily explore insights and interact with the data, ensuring that even

non-technical users can make data-driven decisions. Once the data is processed and structured in the backend, it is sent to the frontend where it is displayed in **real-time** on **dashboards**. These dashboards present key metrics such as **uptime**, **downtime**, **OEE scores**, and **product quality** in an easy-to-understand format. The data is visualized using interactive charts, graphs, and tables that allow users to quickly identify trends, anomalies, and areas for improvement.

The frontend also includes features such as **filters**, **search functions**, and **dynamic reporting tools**, which allow users to drill down into specific data points or time periods for more detailed analysis. These tools make it easier for users to identify root causes of production inefficiencies and monitor

Fig 5. The dashboard of AutoStream Working which is shown to employees

Module 4:

AutoStream focuses on real-time data analysis and insights generation. After the data is acquired and processed in the earlier modules, it is analyzed in this module to extract actionable insights. The backend, built with Node.js and Express.js, performs advanced data analysis, calculating key metrics like Overall Equipment Effectiveness (OEE), uptime, downtime, and product quality.

The system uses real-time processing to ensure that insights are always up to date, providing businesses with accurate performance data as soon as it becomes available. The analysis includes trend identification, anomaly detection, and predictive analytics, helping businesses optimize production processes and improve efficiency.

The backend also integrates with the frontend through APIs, delivering the analyzed data to the React.js-based dashboard where users can view visualizations, charts, and reports that highlight key performance indicators. These insights are presented in an easy-to-understand format allows.

Fig 6. The Dashboard which displays the overall data

Module 5:

Additionally, the module supports alerting mechanisms, where users are notified of any significant issues, such as downtime or quality defects, enabling quick intervention to mitigate potential losses. This real-time analysis ensures that businesses are always aware of their operational performance and can act swiftly to improve their processes.

In summary, Module 5 provides the core data analysis functionality, transforming raw data into actionable insights and ensuring that businesses can continuously monitor and optimize their operations based on real-time information.

Fig 8. MongoDB Storing the data of each employee's OEE

V. SYSTEM ARCHITECTURE

AutoStream is built on a modular, multi-layered architecture designed for efficiency, scalability, and security. Each component of the system plays a distinct role in streamlining the data analysis pipeline from ingestion to insight generation.In the **data acquisition layer**, cameras capture real-time equipment metrics and track SOP compliance. Python, along with libraries like **OpenCV** and **MediaPipe**, monitors hand movements within defined work zones, ensuring actions follow SOP sequences. The data is transmitted via secure APIs to the backend, ensuring continuous data flow for analysis and rapid response.

The **backend processing layer** of AutoStream is built using **Node.js** and **Express.js**, functioning as the core engine for secure data management and system logic. This layer is responsible for computing essential metrics—including operational metrics such as **Availability**, **Performance**, and **Quality**—ensuring that insights are accurate and timely.

It also exposes a set of **secure RESTful APIs**, allowing seamless integration with external enterprise platforms like **Customer Relationship Management (CRM)** systems or other business intelligence tools. These APIs facilitate real-time data synchronization and enhance cross-functional data visibility.

To safeguard sensitive data, the backend enforces robust **authentication and role-based access control**, ensuring that only authorized users—such as operators, maintenance personnel, or managers—can access specific features and data views. By centralizing data processing and access management, this layer guarantees both operationalFinally, the **analytics layer** processes the stored data to generate actionable insights, including real- time analysis, trend detection, and anomaly identification. The system can also trigger alerts for issues such as downtime or quality defects, notifying users immediately for corrective action.

In summary, the **AutoStream system architecture** integrates these five layers to deliver a seamless, efficient, and secure solution for businesses to capture, process, and analyze real-time data, enabling data-driven decision-making and operational improvements.

AutoStream as the data analytics and visualization layer, adds strategic depth to the system by analyzing historical data for trends, production insights, and areas for improvement. This layer pulls data from the backend, transforming raw information into interactive, customizable visualizations on performance trends, SOP violations, and operational bottlenecks. This modular architecture enables the OEE system to operate flexibly and scale across different manufacturing environments, providing continuous performance insights and supporting strategic, data- driven decision-making for long-term efficiency.

DATAFLOW DIAGRAM:

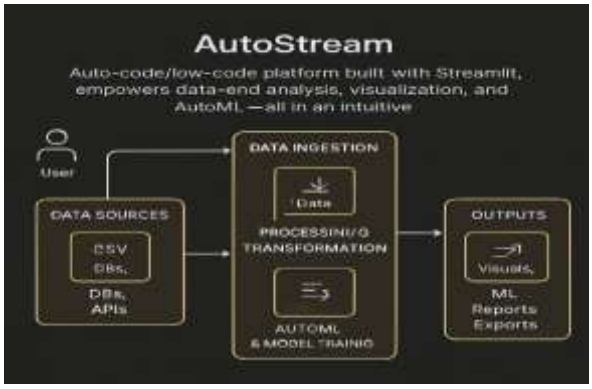

Fig 9. A flow diagram showing the working of the project

II. CONCLUSION

In conclusion, **AutoStream** simplifies the often complex process of data analysis, making it accessible to non-technical users while maintaining the analytical depth required for strategic decision-making. Its ability to handle real-time data processing allows businesses to react swiftly to trends, inefficiencies, or issues as they arise. By unifying data upload, processing, visualization, and model interpretation into a single platform, AutoStream streamlines analytical workflows and significantly reduces time spent on manual tasks.

The platform's focus on ease of use, local-first security, and performance ensures that users can confidently explore and act on their data without relying on multiple disconnected tools. Whether for small teams or larger organizations, AutoStream enhances operational agility and encourages a culture of informed, data-driven decision-making. Ultimately, it equips businesses with the tools they need to remain competitive and continuously improve in an ever-evolving data landscape.

I. REFERENCES

[1] IEEE Xplore: Research papers on AI-driven data analysis, machine learning models for large datasets, and interactive tools for data exploration, relevant for AutoStream's core functionality.

[2] arXiv: A repository for research on conversational AI models, data visualization, and techniques for building user-friendly tools for non- technical users in the context of large datasets.M. Moeuf, R. Pellerin, S. Lamouri, R. Tamayo- Giraldo, and J. Barbaray, "The Industrial Management of SMEs in the Era of Industry 4.0," Computers in Industry, vol. 101, pp. 128-143, Apr. 2018.

[3] "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig: A fundamental text on AI that provides insight into algorithms and methodologies that can inform the intelligent features of AutoStream.A. Gunasekaran, H. Subramanian, and S. Rahman, "IoT in Manufacturing: A Roadmap for Industry Transformation," International Journal of Production Research, vol. 56, no. 18, pp. 5980-5995,
2018.

[4] Data Science for Business" by Foster Provost and Tom Fawcett: A practical guide to data science techniques that can be applied to business problems, relevant for how AutoStream simplifies data exploration for users.A. Solis, B. S. Gonzales, and C. Smith, "Advancements in Predictive Maintenance and Their Impact on Manufacturing Efficiency," Journal of Manufacturing Technology Management, vol. 31, no. 6, pp. 1064-1080, 2020.

[5] Kaggle Datasets: A platform for exploring and downloading diverse datasets to train, test, and enhance the data models behind AutoStream.

[6] Llama3 Model: The large-scale language model used in AutoStream for natural language processing, helping to generate insights and facilitate data-driven conversations with users.

[7] GitHub: A code repository and collaboration platform that hosts open-source projects, including those related to Streamlit, data analysis, and AI, which can help further improve the functionality of AutoStream

[8] OpenAI API: Provides access to advanced AI models for natural language understanding and generation, crucial for the conversational features of AutoStream.