GENERATIVE AND PREDICTIVE AI FOR IMPROVED INDUSTRIAL PROCESSES

An Undergraduate Project Report submitted to Manipal Academy of Higher Education in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Mechatronics Engineering

Submitted by

Aryaman Singh Dev
200929064
Aryaman.dev@learner.manipal.edu
8851759219

Under the guidance of

MR. MAHESH ANIL INAMDAR
Assistant Professor
Department of Mechatronics
MANIPAL INSTITUTE OF TECHNOLOGY

MR. SUDHANSHU KASHYAP
DIRECTOR - AI
SymphonyAI, 1 EPIP Zone,
Whitefield, Bengaluru

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

August, 2024

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

Manipal

12/08/2024

CERTIFICATE

This is to certify that the project titled **GENERATIVE AND PREDICTIVE AI FOR IMPROVED INDUSTRIAL PROCESSES** is a record of the bonafide work done by ARYAMAN SINGH DEV (200929064) submitted in partial fulfillment of the requirements for the award of the degree of BACHELOR OF TECHNOLOGY in MECHATRONICS ENGINEERING of Manipal Institute of Technology, Manipal, Karnataka (A constituent unit of Manipal Academy of Higher Education, Manipal) during the year 2023-2024.

Mr. MAHESH ANIL INAMDAR.
Assistant Professor
Department of MECHATRONICS

Dr. DATTAGURU V. KAMATH
Head of the Department
Department of MECHATRONICS

MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

SymphonyAI

Date: 30-Jun-2024

INTERNSHIP COMPLETION CERTIFICATE

Dear Sir/Madam,

This is to certify that **Aryaman Singh Dev**, a student of **B. Tech** (Branch – MTE) **MIT, Manipal** has completed his internship in the field of **Data Science and Analytics** from 15-Jan-2024 till 30-Jun-2024 at **SymphonyIncubator Business Services Private Limited** under the guidance of **Mr. Sudhanshu Kashyap (Director-AI)**.

During his internship he demonstrated full dedication and self-motivation to learn cutting edge technologies like "Data Cleaning, Exploratory Data Analysis, Visualization, Anomaly Detection ML Model Building, and IRIS Tenant Setup" to name a few. His performance exceeded our expectations, and he was able to offer valuable solutions.

We wish him all the best for his future endeavors.

For **SymphonyIncubator Business Services Private Limited**

DocuSigned by:
Poornima Narayanappa
EC6DBE062C254C3...

Poornima Narayanappa
Senior Manager – Human Resources

SymphonyAI
Tower 3, 6th Floor, SJR Park, EPIP Zone, Whitefield, Bengaluru, Karnataka. 560066

MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to SymphonyAI for offering me the internship opportunity and for their unwavering support throughout its duration.

I am deeply appreciative of Mr. Arjun Bhattacharyya, Senior Vice-President, for sharing his knowledge and wisdom, and for providing an environment where I could work and learn effectively. The years of experience were necessary in providing the drive for the completion of the project.

Special thanks to my mentor, Mr. Sudhanshu Kashyap, Director – AI, for his guidance, encouragement, and invaluable insights. His mentorship has been crucial in developing my understanding of AI-based solutions in the industrial sector. The methodologies were made simpler to understand under his effortless guidance.

Along with that, I am deeply grateful to Mr. Gunaranjan Chaudhary and Mr. Nicolai Nielsen for their unwavering guidance and support for completion of the P&ID tasks. I am also thankful to the entire Energy sector team for their collaborative spirit and willingness to share their expertise. Working under talented individuals has been both enriching and inspiring. Specifically, Mr. Venkatesh Raman, Mr. Vaibhav Malik and Mr. Vaibhav Nadkarni provided invaluable insights for successful completion of the project.

Adding to that, I am obliged to express my gratitude to Mr. Prateek Sharma and Mr. Tushar Bisen for providing a general walk through and implementation for various SymphonyAI platforms and portraying a motivating attitude.

Lastly, I am extremely appreciative to the entire HR team, specifically Mr. Kaustubh Chikhalikar, and the IT team, specifically Mr. Mohan Senthil for their unwavering support to ease the ingestion and completion phases.

ABSTRACT

This report delves into the utilization of Generative AI (GenAI) and Predictive AI (PredAI) to address various problem statements in the industrial sector. The primary issues identified include operational inefficiencies, high maintenance costs, quality control challenges, decision-making complexities, and the need for adaptability to market changes. By leveraging AI technologies, these challenges are met with innovative solutions that streamline processes, enhance decision-making, and optimize operations.

The manufacturing industry has long grappled with inefficiencies, quality control issues, high maintenance costs, and supply chain management challenges. These problems manifest in various forms, including repetitive labor-intensive tasks, equipment failures, inconsistent product quality, and inaccurate demand forecasting. These issues necessitate the integration of advanced AI technologies to improve operational efficiency, reduce costs, and maintain competitiveness. GenAI enhances manufacturing reliability and process optimization by analyzing plant performance to identify trends, patterns, and anomalies. It also supports digital manufacturing and helps operators enhance efficiency by preventing downtime through large language models (LLMs) and AI copilots. PredAI uses data analysis and machine learning to forecast future events and trends. In industrial automation, PredAI improves decision-making processes and operational efficiency by predicting machinery malfunctions, supply chain disruptions, and market demands. It also plays a crucial role in quality control by detecting anomalies and defects early, thus maintaining high standards and reducing waste.

The integration of GenAI and PrediAI has resulted in several positive outcomes discussed below. Automation of repetitive tasks allows human workers to focus on strategic and creative activities, thereby increasing productivity, job satisfaction, and efficiency. Predictive maintenance specifically reduces downtime and maintenance costs by forecasting potential equipment failures and allowing timely interventions. Moreover, AI-powered systems ensure higher accuracy and consistency in product quality by inspecting products in real time, thereby offering improved quality control. Along with that, accurate demand forecasting and inventory optimization reduce costs and improve customer satisfaction to optimize supply chain management. In addition to the above-mentioned positives, AI-driven decision support systems also provide real-time data analysis, aiding managers in making better production schedules, resource allocation, and process optimization decisions. This leads to a series of informed decision making. Lastly, predictive insights enable manufacturers to quickly adapt to market fluctuations and disruptions, enhancing agility.

Generative AI and Predictive AI are widely utilized in various industrial applications. Continuous monitoring and analysis of production lines to detect and address defects early optimizes quality control. Forecasting potential equipment failures to reduce downtime and maintenance costs aiding predictive maintenance. Supply chain management can optimize inventory levels and demand forecasting to improve efficiency and reduce costs. Finally, operational optimization can be achieved by enhancing process efficiency and reducing labor-intensive tasks through automation

The primary goal was to enhance operational efficiency by detecting bottlenecks and anomalies to reduce costs and improve product quality through the integration of AI technologies. These goals were achieved through the deployment of Generative AI and Predictive AI systems, resulting in improved decision-making, optimized processes, and increased adaptability to market changes. The outcomes highlight the significant potential of AI-driven solutions in transforming the industrial sector and addressing its longstanding challenges.

SymphonyAI's industrial AI applications, such as IRIS Workspace, Predictive Asset Intelligence, Enterprise KPI dashboard and Vision AI, exemplify the practical integration of GenAI and PredAI. These tools enable real-time monitoring, predictive maintenance, and enhanced quality control, thereby improving operational efficiency and reducing costs. SymphonyAI’s comprehensive AI solutions offer rapid deployment and scalability, ensuring swift returns on investment and continuous operational improvement across various industrial domains. In conclusion, the application of GenAI and PredAI in the industrial sector has demonstrated substantial improvements in efficiency, cost reduction, and product quality. These technologies provide a robust framework for addressing complex industrial challenges and achieving sustainable growth and competitiveness.

LIST OF FIGURES

Table 1

CONTENTS

CHAPTER 1

INTRODUCTION

1.1) General Discussion

Data Scientists leverage their knowledge to accelerate revenue growth and enhance business through algorithmic building. Developments in the fields of (Artificial Intelligence) AI, both (generative AI) GenAI and (predictive AI) PredAI has enhanced decision-making, which has resulted in lower wastage in comparison.

The culmination of AI with industrial processes has developed remarkably in recent years allowing for faster decision-making by data-driven insights. Proceedix, a connected worker end-to-end integration allows third party integrations with external API's (Application Programming Interface) and ERP (Enterprise Resource Planning) Systems. Machine learning algorithms identify customers' buying habits by understanding historic trends to analyse the data. Platforms like (Anomaly Detection Engine) ADE and MLOps (Machine Learning Operations) Studio were used to work around the (Key Performance Indicator) KPI's and Knowledge graphs were visualized through the platform to work on the areas of improvement.

1.2) Present-Day Scenario

AI-powered smart manufacturing solutions drive revenue generation through optimized operations. GenAI and PredAI softwares improve manufacturing analytics, copilots solution, and performance monitoring. The IRIS (Industry Reasoning and insights Service) Workspace confidently makes data-driven decisions in the form of time-series analysis and visualized analysis. MLOps and industrial copilots predict and detect issues in performance. ADE is used to detect anomalies .

Fig. 1-SymphonyAI

1.2.1) AI-Driven Manufacturing

1. Optimizing Operations and Asset Health

Asset health can be tracked by an evaluated health score to help optimizing the operations and maximize the revenue by generating a higher ROI (Return on Investment).

2. Manufacturing Operations Management and Analytics

Predictive maintenance schedules repairs during quieter hours to reduce operational disruptions and costs.

Real-time analytics constantly monitor data to improve production, minimize defects and ensuring consistent product quality.

3. Manufacturing Copilots

Copilots helps teams collaborate more effectively and provide solutions in case of any errors.

1.2.2) Broader Impact of AI in Industrial Automation

1. Enhanced Agility and Quality

AI enables manufacturers to adapt to market changes. Tracking historical trends helps predict and generate futuristic requirements beforehand, thus cutting on waste.

2. Informed Decision-Making

Visualized patterns help identify faults and defects, like ADE. Through machine learning algorithms, a data can be presented in a formatted way to find the bottleneck.

3. AI in ERP Systems

Improved data processing & analytics is transformative for ERP systems and offering better forecasting.

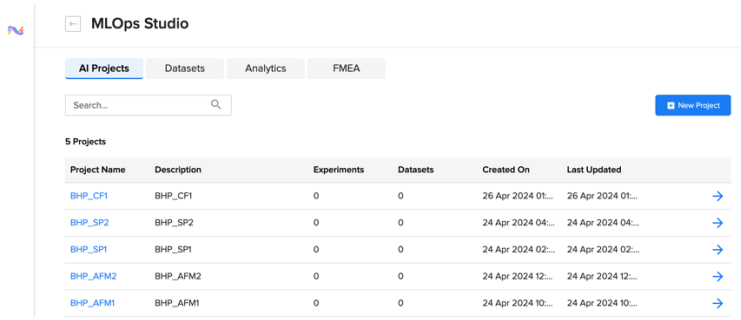

Fig. 2-MLOps Studio

1.3) Motivation to do the Project Work

Researching about the impact of AI on society today offered invaluable information and made me keen to pursue this field in the future. This project targets at utilizing a structured methodology to harness gaining knowledge about the pipeline of an AI project in the manufacturing industry rather than working on an a single project. It has presented a better way to learn and practice on the same step before going to the next step in the methodology.

The inclusion of AI in industrial automation can be used to improve efficiency, reduce costs, and thereby increase productivity.

1.3.1) Utilizing AI in industrial automation

1. Predictive Maintenance

Equipment performance metrics were measured and visualized to the platform. Historical trends were utilized to forecast potential machine or component failures. Patterns were drawn as a conclusion to aim at predictive maintenance thereby reducing wastes.

2. Quality Control and Inspection

The P&ID project focused on implementing the YOLO model for tags and assets recognition to ensure the correctness of the diagram before utilizing machine learning algorithms. This allowed for anomaly detection due to wrong naming or improper detection, and mainly focused on manually checking for errors before allowing the model to detect by itself. This functionality focused on a double-check to retrain the model on the incorrect detections. This facilitates real-time quality control in the long term.

3. Process Optimization

By discerning patterns, correlations, and inefficiencies within the data, AI systems can optimize process parameters, forecast optimal production schedules, and recommend enhancements to overall manufacturing operations. This was impacted with the help of manual inspection of the visualized charts after data cleaning to fulfil the tasks before model training.

4. Demand Forecasting

Helped reduce inventory carrying cost by analysing patterns in sales based on seasons, or holidays to more accurately predict future demand.

5. Decision Support Systems

GenAI and PredAI platforms focused on acting real-time data analysis to generate insights, and offering recommendations. This was necessary to find the bottleneck causing the inefficiency.

7. Personalized solutions

The work included understanding different data from different clients and offering a solution to the individualized problems. Mostly ensured by cleaning and visualizing the data in the form of boxplots or time-series analysis to project potential future by analysing patterns to detect anomalies, eventually improving the performance.

MLOps Studio in Workbench and Energy and Petrochemical predictive portals ensured the attainment of these procedures.

1.4) Shortcomings in Previous Work

- Accessibility to industrial data when building, deploying and scaling industrial solutions.
- Managing the rapidly increasing data generation at industrial scale.
- Industrial data lacks context.

- Slow moving digital initiatives, unable to scale beyond one-time use cases.
- PredAI models require tedious manual processes.
- GenAI chatbots are limited to search and code.
- Manufacturing workflows remain largely unchanged, requiring pre-built industrial applications for data-driven decisions.
- Inadequate tools for EDA (Exploratory Data Analysis).
- Unable to accurately measure demand forecasting.
- Need for real-time analysis to improve decision making and optimize performance.
- Insufficient detection of defects for vision-based models.

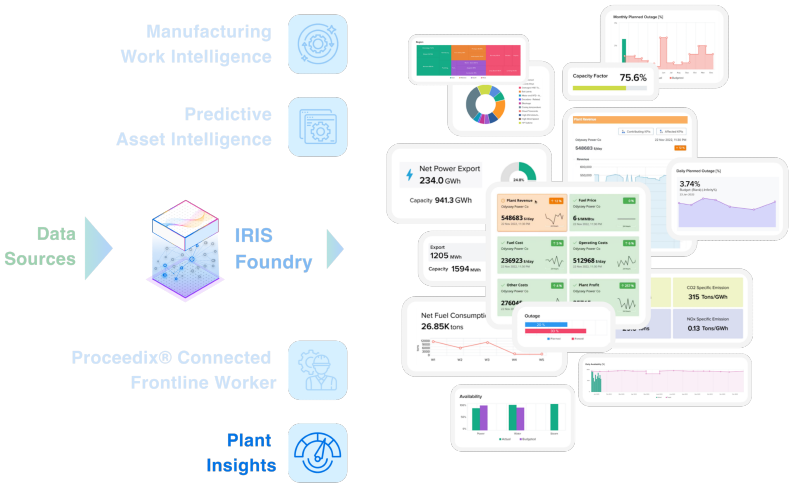

Fig. 3-generate actionable KPIs with Plant Insights [1]

1.5) Importance of the Work in the Present Context

In the current context, this work is crucial as it addresses the need for industry-specific AI applications that can deliver real-time insights and optimize operations efficiently. The project focusing on Energy and Petrochemical tenants revolved around presenting the health score and critical assets severity based on current data and a series of historical trends to focus on predictive futuristic analysis.

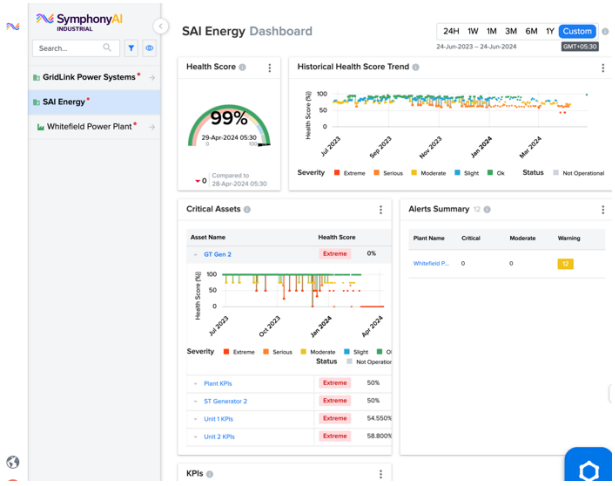

Fig. 4-Energy predictive portal dashboard

Groundbreaking Industrial Large Language Model (LLM) with the help of GenAI enhances manufacturing, and process optimization to help with faster decision making.

The Industrial LLM trained on 1.2 billion tokens using one of the largest industrial datasets, comprising of 3 trillion data points, more than 500,000 machine tests, 150,000 components and 80,000 different assets improves operational efficiency, productivity and margins. The LLM was useful for providing context-aware data and delivers actionable knowledge to operators up to 90% more rapidly than their existing systems. Thus, leveraging and building solutions harbouring personalized data not only offers efficient results for the above problems but also enhances the ROI by cutting costs.

Along with this, Industrial Reasoning and Insights Service (IRIS) Foundry, an Industrial DataOps platform can rapidly create robust predictive and generative AI industrial applications to reduce unscheduled asset downtime, improve process efficiency, and enhance connected worker operations. It helps in transforming the way industries manage their data with the help of real-time AI-powered insights.

Automated P&ID Ingestion in IRIS allows interpreting drawings and processing patterns with vision and ML algorithms to identify assets, tags and connections. Real-time KPI monitoring allows for reduced maintenance-time and waste.

Energy and Petrochemical plant insights allowed automated data pre-processing and MLOps pipeline effectively implemented. Kubeflow portal allowed accessible solutions and structured data for every individual project in hand.

1.6) Uniqueness of the Methodology

- The methodology adopted in this project is unique due to its focus on Industrial LLMs and combining them with knowledge graphs to create personalized AI solutions.

- The utilization of the IRIS MLOps Studio Platform and Workbench for both Energy and Petrochemical tenants enhances the process efficiency by showcasing the health score for various assets.

- Historical trends studied in PredAI and GenAI to get a layout for futuristic events.

- Allows enhancing proficiency and understanding the bottlenecks with visualization.

1.7) Significance of the Possible End Result

The IRIS Foundry enhances predictive analytics by allowing seamless data orchestration like P&ID ingestor and provide real-time monitoring to scale built-in models. This allowed integration with manufacturers to create robust predictive and generative AI solutions to improve process efficiency, cut down costs and unscheduled asset downtime, as well as enhance the connected worker operations.

The rapid and relevant results included, but are not limited to:

- 2-4% increase in product throughput
- 20-50% reduction in unplanned downtime
- 3-5% increase in energy efficiency

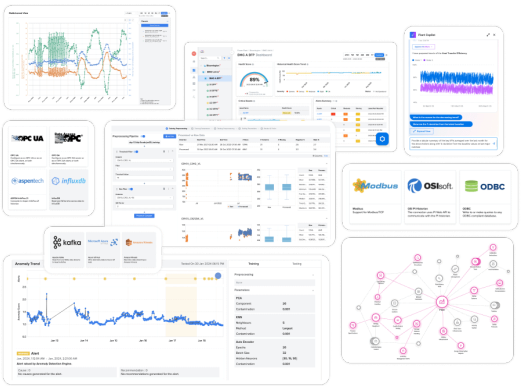

Fig. 5-Industrial AI software applications [2]

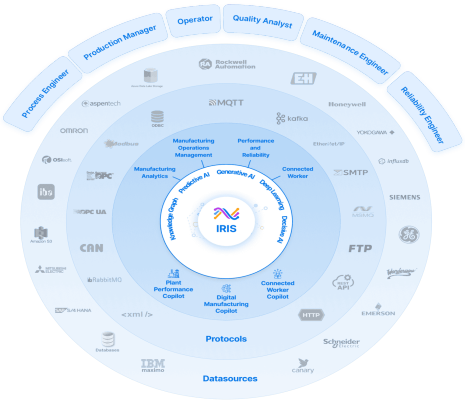

Fig. 6-Industrial Connectivity [3]

CHAPTER 2

LITERATURE REVIEW

2.1) Present Day State in industrial automation

The collaboration of industry experts with leading AI technology has resulted in uniquely useful applications. Vertical and Horizontal domain expertise allows the integration of solutions that reduce the time to value ratio. Immediate benefits include real-time insights and analytical optimization, portraying the health score to increase the asset lifetime and ROI.

2.1.1) Generative AI in Industrial Automation

Use of algorithms to create new content from the existing content, which is particularly beneficial in manufacturing and product development using prompts trained on LLMs. In industrial settings, GenAI is employed to optimize design processes and develop new prototypes rapidly. This capability allows companies to reduce the time and costs associated with traditional design and testing phases.

One notable application of GenAI is in the customization of manufacturing processes. By generating multiple design variations based on predefined parameters, GenAI helps identify the most efficient and effective solutions. This technology is also instrumental in predictive maintenance, where it generates models that can foresee potential equipment failures, thereby reducing downtime and maintenance costs.

2.1.2) Predictive AI in Industrial Automation

Predictive AI, on the other hand, uses data analysis and machine learning to forecast future events and trends. In industrial automation, PredAI is crucial for enhancing decision-making processes and improving operational efficiency. By analyzing historical data and identifying patterns, PredAI can predict machinery malfunctions, supply chain disruptions, and market demands.

A significant application of PredAI is in quality control. By continuously monitoring production lines and analyzing data in real time, PredAI systems can detect anomalies and defects early.

This proactive approach ensures that potential issues are addressed before they escalate, thereby maintaining high-quality standards and reducing waste. Additionally, PredAI can optimize inventory management by predicting demand fluctuations and adjusting supply levels accordingly, which minimizes overproduction and stockouts.

2.1.3) The Impact on the Workforce

The integration of GenAI and PredAI into industrial automation is significantly altering the workforce landscape. Allowing AI technologies to perform repetitive, labor-intensive tasks, allows human workers to focus on more strategic, creative, and problem-solving activities. This shift not only enhances productivity but also improves job satisfaction by engaging workers in more meaningful work. This approach fosters a collaborative environment where human expertise and AI capabilities complement each other, driving innovation and efficiency.

2.1.4) Case Studies in SymphonyAI

1. Industrial LLM

   Allows the correlation of manufacturing problem solving, data analysis, and reporting with Generative AI by enhancing manufacturing, reliability, and process optimization. Performs analysis on plant performance to identify trends, patterns and anomalies, digital manufacturing to assist operators to enhance efficiency and help the connected workers prevent downtime using LLM.

   By permitting independent operation through an API, chatbot access, or industrial copilots, it presents a seamless integration of real-time data sources with knowledge graphs.

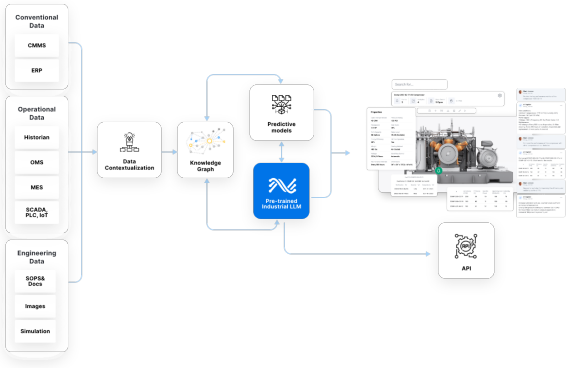

Fig. 7-Industrial LLM [2]

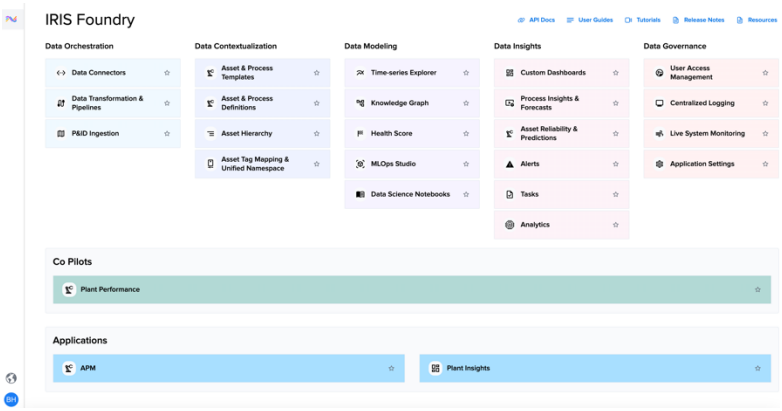

Fig. 8-IRIS Foundry

2. IRIS Workspace

Proposes a live collaboration environment to analyze and troubleshoot industrial operations. Powered by dataops and generative AI copilot, it transitions empowering cross-functional teams to engage in time-series analysis, P&IDs (piping and instrumentation diagrams), documentation and drawings catered to solving problems.

It seamlessly allows version tracking, pattern recognition, and recommendations using generative and predictive AI to generate real-time insights on industrial data.

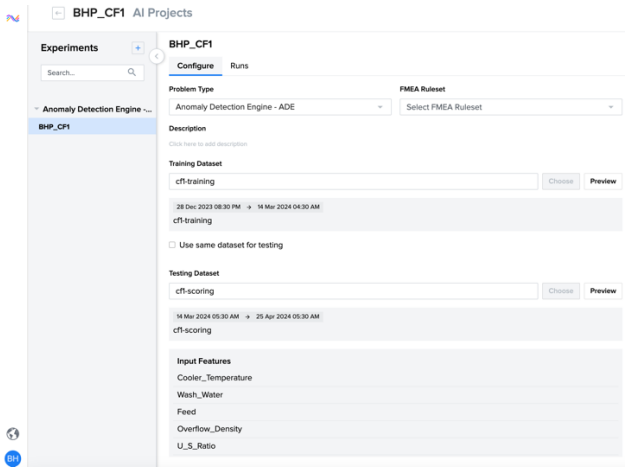

Fig. 9-MLOps Studio Centrifugal

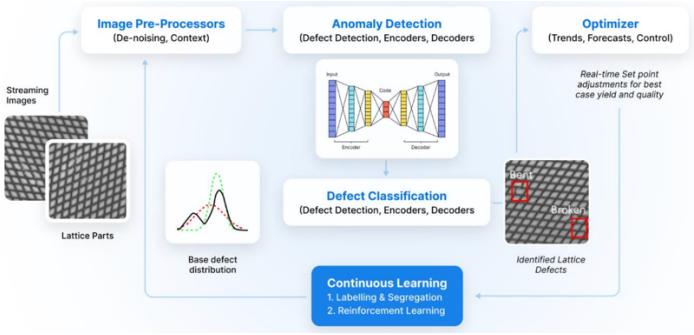

Fig. 10-IRIS pipeline [4]

3. Eureka GenAI platform

An industry leading multi-model architecture allowing rapid application development by utilizing copilots. Consisting of state-of-the-art Lakehouse architecture and vertical specific LLMs, GenAI skills and LLM-based orchestration, along with NLP integration in UI/UX it solves the problem of rapid development of applications.

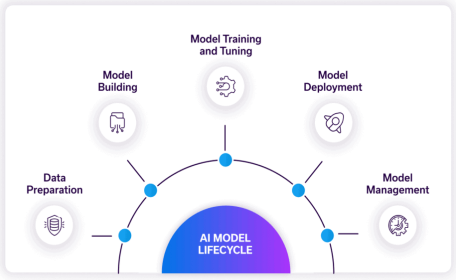

Fig. 11-Lifecycle Management [5]

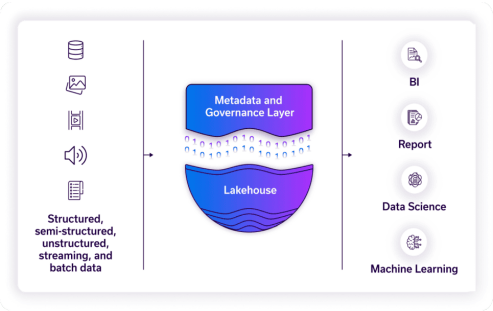

Fig. 12-Lakehouse Architecture [5]

The customized AI copilots understand the challenges fueled by GenAI and PredAI to optimize ML lifecycles for individual vertical applications. The lakehouse architecture allows accessing structured and unstructured data and the vertical LLMs contextualize user queries for each individual case.

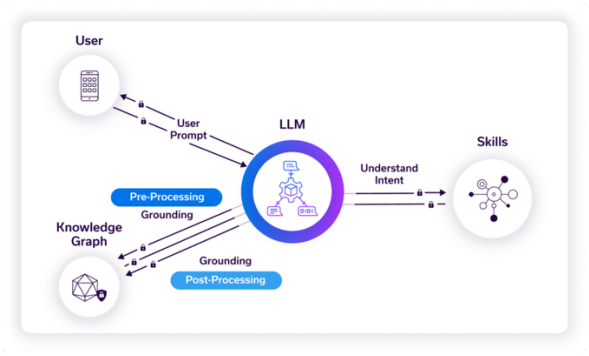

Fig. 13- Vertical LLMs structure [5]

4. IRIS Vision AI and Proccedix

   Vision AI studio automates manufacturing workflows by identifying anomalies and deviations using time series analysis and work orders.

Fig. 14-Vision AI [4]

Proceedix offers smoother inspections, reduced reworks, improved process excellence, and full traceability to digitize instructions and inspections using tablets, desktops, or smart glasses.

CHAPTER 3

PROBLEM DEFINITION

The main problems in the issue of managing data in the manufacturing industry has been looming for a long time leading to quality control issues, poor maintenance and waste, lack of continuous improvement and process optimization, and inaccurate decision making. This along with poor real-time performance analysis, lack of demand forecasting and failure to recognize the trend led to inclusion of AI to solve industrial issues, some of which are listed below:

1. Operational Inefficiencies

Repetitive, labour-intensive tasks slowing down production, need to be automated to streamline operations.

2. High Maintenance Costs and Downtime

Adhering to equipment failures and unplanned downtime well in time can aid the growth of the company by controlling the manufacturing operations, and substantial financial losses. Predictive maintenance powered by AI algorithms helps in forecasting potential failures, allowing for timely interventions and reducing downtime and maintenance costs.

3. Quality Control Issues

Maintaining consistent product quality is critical in manufacturing. Manual inspection processes are prone to errors and inefficiencies. AI-powered quality control systems can inspect products in real-time, ensuring higher accuracy and consistency. The P&ID systems utilizing YOLO ensured completion of manual inspection errors before pairing it with technology for double tracking of errors.

4. Decision-Making Complexities

Lack of real-time data analysis has affected the decision-making abilities. Helping managers make better decisions regarding production schedules, resource allocation, and process optimization can help in cutting costs.

5. **Adaptability to Market Changes**

   Market agility enhances the ability of growth by offering new factors in the spectrum of possibilities. Real-time tracking allows continuous growth and adaptability.

6. **Automated P&ID ingestion**

   P&ID integration using computer vision models to enhance the detection and correction of errors to establish an effective solution.

7. **Implementing Predictive maintenance**

   Deploying PredAI and GenAI at scale without tedious manual inspection processes to explore and clean data and implementing faster model building.

8. **Increasing data volume**

   Exponential increase in data generation at industrial scale and issue of addressing and managing the ever-growing amount of data.

9. **Early-Warning System**

   Utilizing multi-model ADE seamlessly integrates data, documents and images and provide timely performance problems.

By focusing on these issues, teams can collaborate together to inculcate AI integration as a mandatory requirement to enhance efficiency, reduce costs, and maintain competitiveness. Various SymphonyAI solutions can cater to these problems, some of them like computer vision using YOLO and P&ID ingestor, copilots and industrial solution platforms to streamline these processes and present a financially efficient solution.

CHAPTER 4

OBJECTIVES

1. Industrial Automation

- Objective: Optimize manufacturing operations and efficiency
- Project Focus: Work on AI-powered manufacturing analytics to provide real-time plant performance insights and improve operational efficiency by finding anomalies.
- Specific Goals:
  - Implement manufacturing copilots leveraging generative AI to support operational decision-making.

  - Enhance performance and reliability monitoring systems for better asset management.

  - Empower operations excellence through integrated AI applications that streamline manufacturing processes.

2. Cross-Domain AI Solutions

- Objective: Develop and deploy generative and predictive AI solutions to address specific industry needs.
- Project Focus: Utilize the MLOps Studio and IRIS Foundry to create scalable, integrative AI solutions.
- Specific Goals:
  - Fine-tune large language models (LLMs) and integrate with knowledge graphs to build industry-specific RAGs.
  - Enhance process efficiency and minimize downtime with Data Ops platforms.
  - Integrate AI solutions with existing workflows to provide seamless and impactful business applications.

3. Industry knowledge Objectives

- Understand the P&ID assets, tags and endpoint connections.
- Use electronics domain knowledge to understand the problem statement before diving deeper into working on the data.
- Implement P&ID diagrams from scratch

4. Data Scientist Objectives

- Read excel file data to analyse and clean
- Aim at developing high-accuracy predictive models tailored to specific industry needs. Using historical data to predict future trends and behaviours in industry and retail sectors
- Enhance data processing and cleaning pipelines to ensure high-quality data inputs.
- Implement robust data preprocessing steps to clean, structure, and validate data before analysis.
- Create and maintain comprehensive documentation for all data science processes and models.
- Produce detailed documentation covering data sources, preprocessing steps, modelling techniques, and evaluation metrics.
- Facilitate effective collaboration between data science and other departments by understanding electronics and hardware in industries.
- Implement advanced machine learning and AI techniques to solve complex business problems.
- Monitor and refine deployed models to ensure continuous improvement and relevance.
- Track the performance of models in production and make necessary adjustments to maintain high standards.
- Research about Yolo and other computer vision models.
- Contribute to the development of the MLOps DTS Studio and IRIS Foundry and enhance the capabilities of SymphonyAI's AI platforms to support diverse industry applications.

CHAPTER 5

THEORETICAL BACKGROUND

5.1) Theoretical Discussions and Background

The need for efficient solutions at the industrial level ensured thorough practicality and completion of the projects to promote system orchestration, and promising enhanced productivity whilst redefining creativity.

5.1.1) Computer Vision

Utilized for evaluating misclassifications in P&ID diagrams. Theoretical basis involves convolutional neural networks (CNNs) and transfer learning, which enable pre-trained models to be adapted for specific tasks.

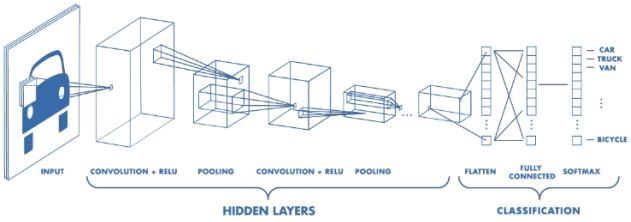

Fig. 15-CNN architecture [6]

Fig. 16-JSON format

5.1.2) JSON File Handling and Data Annotation

Critical for preparing datasets representative of a data structure as text files for ease of storage and transmitting information across a network.

5.1.3) Failure Mode and Effect Analysis (FMEA)

A systematic method for evaluating processes to identify potential failure points. This involves statistical analysis and risk assessment techniques.

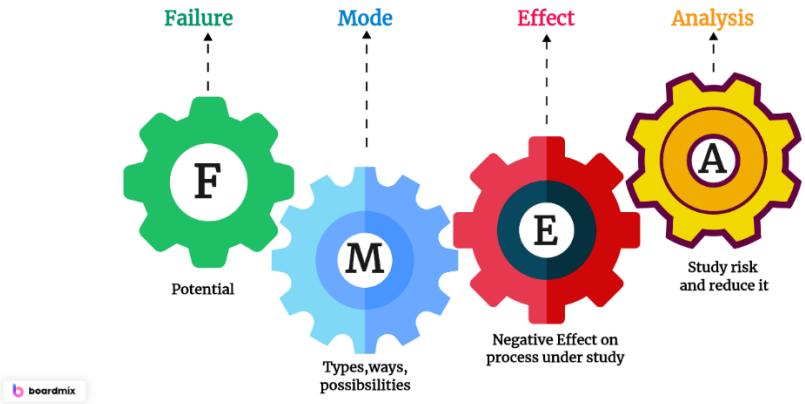

Fig. 17-FMEA analysis [7]

5.1.4) Risk Priority Number (RPN)

$$RPN = Severity \times Occurrence \times Detection$$

Each factor is rated on a scale, and the RPN helps prioritize risks based on their criticality.

		Consequence				
		Insignificant	Minor	Moderate	Major	Severe
Likelihood	Almost certain	Medium	High	High	Extreme	Extreme
	Likely	Medium	Medium	High	Extreme	Extreme
	Possible	Low	Medium	Medium	High	Extreme
	Unlikely	Low	Low	Medium	High	High
	Rare	Low	Low	Low	Medium	High

Fig. 18-FMEA chart [8]

5.1.5) Anomaly Detection Techniques

Implemented using PCA (Principal Component Analysis), KNN (K-Nearest Neighbors), and Autoencoders. These methods are grounded in statistical analysis, dimensionality reduction, and neural networks, respectively.

5.1.5.1) Principal Component Analysis (PCA)

PCA transformation reduces dimensionality by projecting data onto principal components.

$$Z = X . W$$

$$X = Z . W \hat{} T$$

1. Standardize Data: Normalize the dataset to have a mean of zero and standard deviation of one.

2. Compute Covariance Matrix: Calculate the covariance matrix to understand the data's variance structure.

3. Eigen Decomposition: Find the eigenvalues and eigenvectors of the covariance matrix.

4. Sort Eigenvalues: Arrange eigenvalues (and corresponding eigenvectors) in descending order.

5. Select Principal Components: Choose the top k eigenvectors as principal components based on the largest eigenvalues.

6. Transform Data: Project the original data onto the selected principal components to form a reduced dataset.

Fig. 19-PCA methodology

where X is the data matrix, and W are the eigenvectors of the covariance matrix of X.

5.1.5.2) K-Nearest Neighbors (KNN)

KNN identifies anomalies based on distance metrics. Anomalies are points that have fewer than k neighbors within a specified distance.

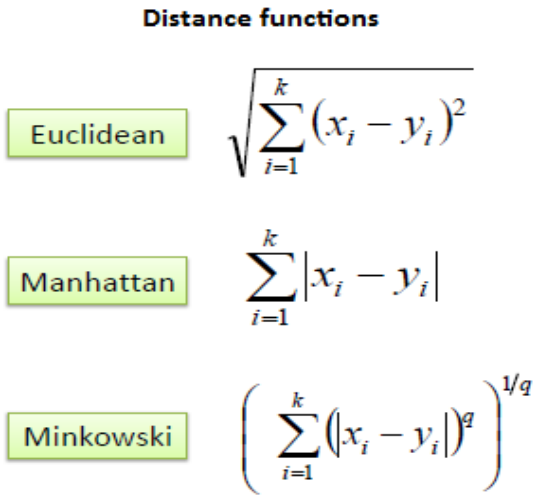

Fig. 20-Distance Metrics [9]

5.1.5.3) Autoencoders

Autoencoders are neural networks designed to learn a compressed representation of data.

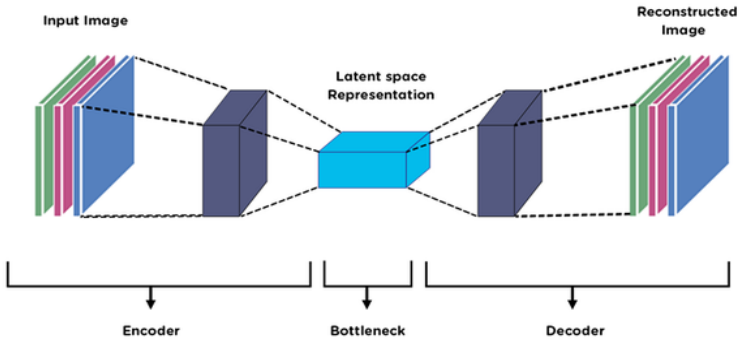

Fig. 21- Autoencoder architecture [10]

CHAPTER 6

METHODOLOGY

6.1) Initial Setup and Research

Focused on acclimatization, understanding SymphonyAI's objectives, and researching SymphonyAI's solutions. Revisited Python libraries and studied Piping and Instrumentation Diagrams (P&ID). Finally, used a pre-built computer vision model to evaluate misclassifications in P&ID diagrams.

6.1.1) Assumptions

- Assumed basic familiarity with Python programming language and its general libraries.
- Assumed access to SymphonyAI's resources and documentation for initial research.
- General understanding on Computer Vision models and techniques.

6.1.2) Design Analysis and Experimental Work

➢ Objective

Understand SymphonyAI's objectives and solutions, revisit Python libraries, study Piping and Instrumentation Diagrams (P&ID), and work with computer vision model to separate misclassifications and correctly classified.

➢ Approach

- Conducted a thorough review of SymphonyAI's documentation and solutions.
- Revisited Python libraries NumPy, pandas, Matplotlib and Seaborn to refresh foundational skills.
- Studied P&ID diagrams to understand their structure and components.
- Understood the connections, assets and tags in the diagram.

6.1.3) Tools and Libraries Used

- Python: NumPy, pandas, Matplotlib, Seaborn
- SymphonyAI documentation and resources
- Computer Vision: YOLO model

6.1.4) Preliminary Results

- Gained a solid understanding of SymphonyAI's objectives and solutions.
- Refreshed knowledge of essential Python libraries.
- Acquired basic knowledge of P&ID diagrams and connections.
- Understood the correctly classified and misclassified data in P&ID diagrams.

6.1.5) Eventual Results and Conclusions

- Successfully completed the initial setup and research phase.
- Established a strong foundation for further work on P&ID diagrams and related tasks.
- Understood general working of computer vision in industrial manufacturing domain.

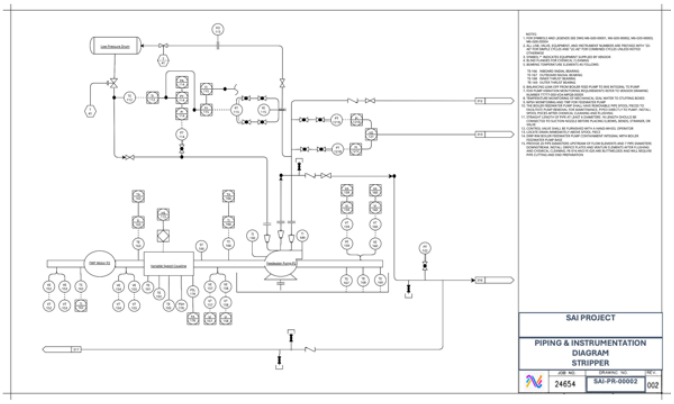

Fig. 22-P&ID diagram

6.2) Deepening Understanding of P&ID Diagrams and JSON Files

Worked with JSON files in Python, checked tag names, ID types, and bounding boxes. Documented errors and misclassifications manually edited JSON files and generated new batch results with fewer errors. Lastly, documented the findings.

6.2.1) Assumptions

- Assumed availability of JSON files containing P&ID annotations.
- Assumed the knowledge for manual inspection and correction of data.
- Assumed the knowledge for Yaml file generation through python, and integration in Postman.

6.2.2) Design Analysis and Experimental Work

➢ Objective

Deepen understanding of P&ID diagrams, work with JSON files in Python, and document errors and misclassifications. Research about algorithms for understanding.

➢ Approach

- Analyzed JSON files to understand their structure and content.
- Verified tag names, ID types, and bounding boxes in P&ID annotations.
- Document the connection errors and misclassifications in JSON files.
- Algorithm revision.

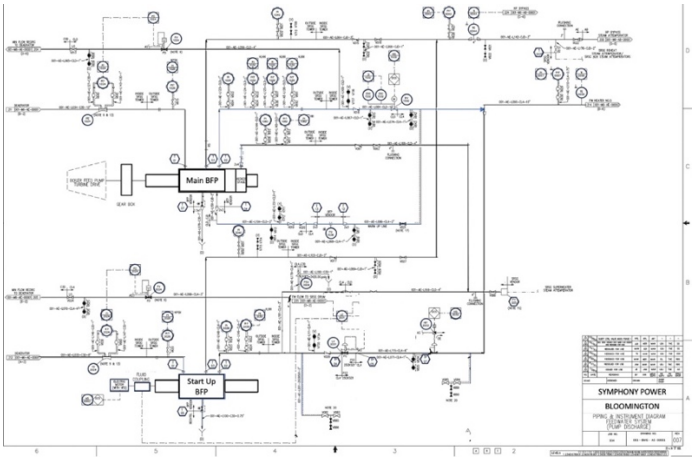

Fig. 23-P&ID chart

6.2.3) Tools and Libraries Used

- Python: json and yaml files, NumPy, Pandas
- Text editors for manual JSON file editing

6.2.4) Preliminary Results

- Identified common errors and misclassifications in P&ID annotations.
- Documented the types and frequencies of these errors.

6.2.5) Eventual Results and Conclusions

- Successfully corrected a significant number of errors and misclassifications.
- Improved the overall quality of P&ID annotations, leading to more accurate analyses.

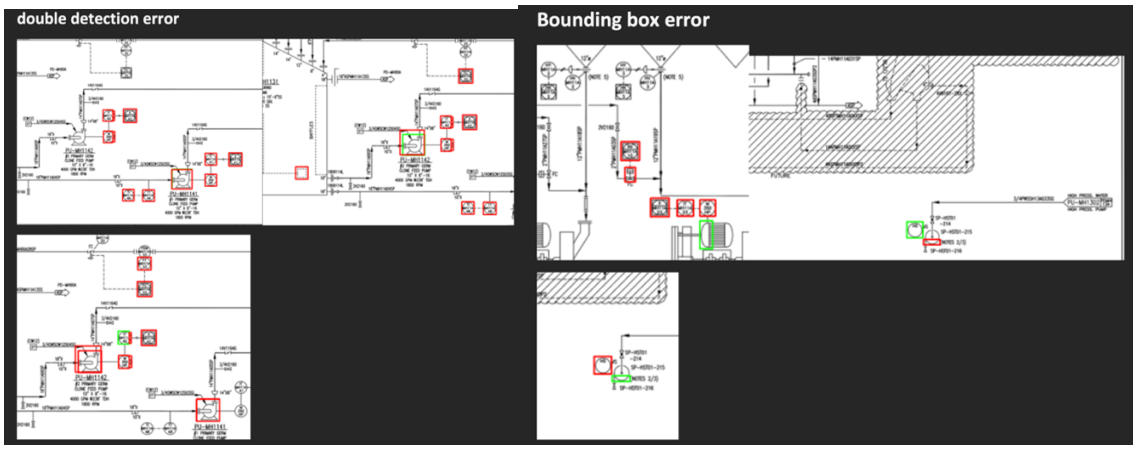

Fig. 24 -Errors

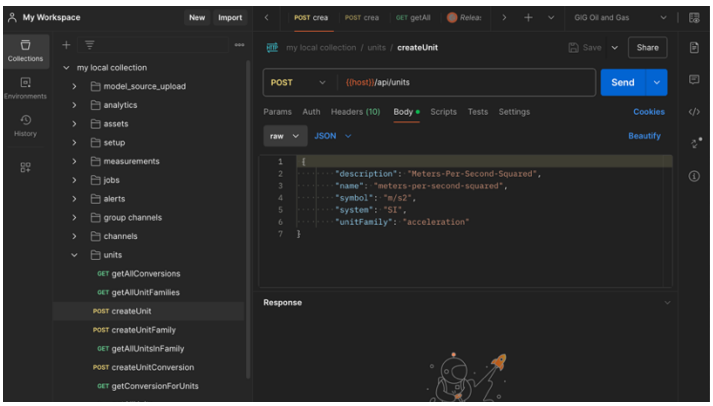

Fig. 25-Postman layout

6.3) Platform Knowledge Expansion and Advanced Models

Learned about Docker, Postman, YOLO models for object detection, and Huggingface Transformers. Created P&ID diagrams, worked on data anonymization and resampling. Along with this, work included researching SymphonyAI's predictive portal, AI studio, APIs in Postman, and fine-tuning a pre-trained model from Huggingface and evaluating YOLO models' performance.

6.3.1) Assumptions

- Assumed access to various platforms and tools; Docker, Postman, YOLO, Huggingface.
- Assumed familiarity with basic concepts of machine learning and deep learning.
- Assumed access to Atlassian collaboration framework at SymphonyAI.
- Credentials to check out the solutions at SymphonyAI – APM 360 Predictive portal, DTS AI Studio, and Eureka Workbench.
- Visual Paradigm software credentials for making P&ID diagrams.
- Access to energy and petrochemical tenants' data.

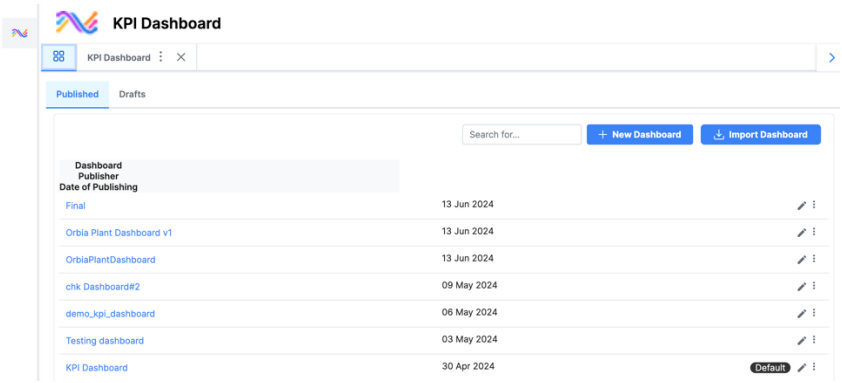

Fig. 26- DTS AI studio

6.3.2) Design Analysis and Experimental Work

➢ Objective

- Evaluate the errors in current P&ID diagram detection.
- Create units for various tags and test on Postman.

- Expand knowledge to research about current solutions of SymphonyAI – APM 360 Predictive portal, DTS AI Studio, and Eureka Workbench.
- Work on Energy and Petrochemical tenants' solutions.
- Docker, Postman, YOLO models, and Huggingface Transformers, and apply them to P&ID diagrams and related tasks.
- Perform data anonymization by resampling timestamp data or changing names for tags.

➢ Approach

- Evaluated the P&ID diagrams after error documentation.
- Learned the basics of Docker and Postman for containerization and API testing.
- Studied YOLO models for real-time object detection and Huggingface Transformers for NLP tasks.
- Created P&ID diagrams, worked on data anonymization, and fine-tuned a pre-trained Huggingface model for learning purposes.
- Calculated the statistics for data using pandas library and visualized the data with matplotlib and seaborn libraries.
- Performed resampling of timestamp for data anonymizing task and plotting time-series analysis.

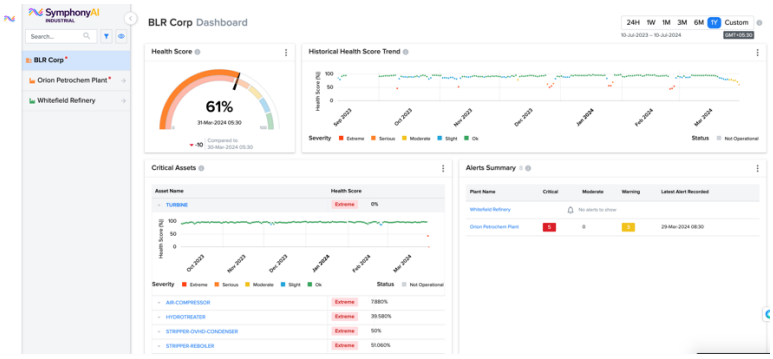

Fig. 27- Refining analytics

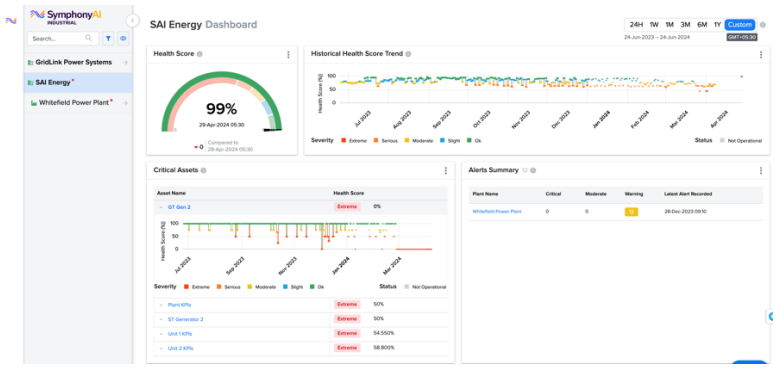

Fig. 28- Energy analytics

6.3.3) *Tools and Libraries Used*

- Docker, Postman, Workbench
- Python: YOLO, Huggingface Transformers
- Pandas for data manipulation
- Numpy to store data and work with lists, arrays and matrices
- Matplotlib and seaborn for visualized analysis

6.3.4) *Preliminary Results*

- Understood Docker containers and tested APIs using Postman.
- Understood fine-tuning Huggingface models for specific NLP tasks.
- Evaluated YOLO models' performance on object detection in P&ID diagrams.
- Performed anonymization on data, both on tags and timestamps and calculated statistics for analysis.

6.3.5) *Eventual Results and Conclusions*

- Enhanced proficiency with Postman testing.
- Created units on Postman for testing.
- Successfully applied these tools to relevant tasks, improving efficiency and accuracy in P&ID analysis and data processing.
- Analyzed data in form on visualized charts like timeseries analysis or boxplots visualization.
- Created P&ID diagrams for energy and petrochemical tenants.

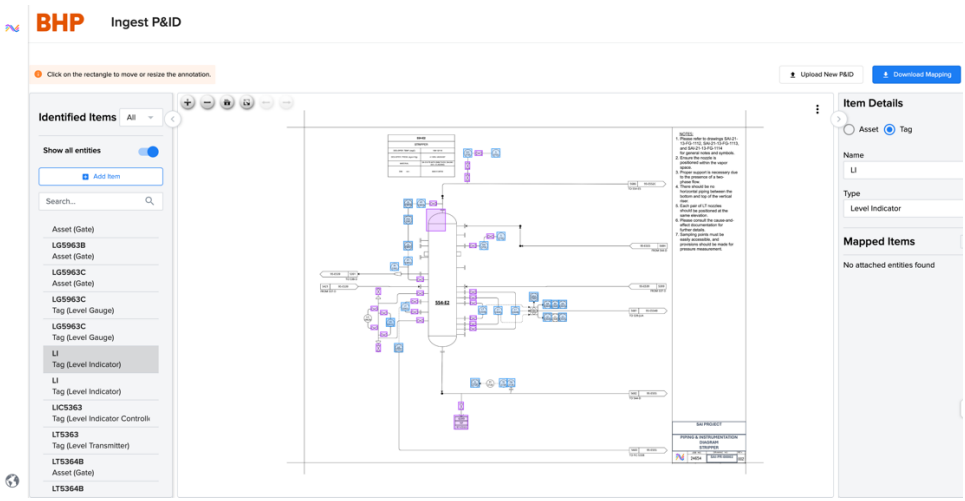

Fig. 29- P&ID ingestor

6.4) Finalizing P&ID Tasks and Advanced Analysis

Explored Failure Mode and Effect Analysis (FMEA) templates, cleaned data, created channels and units on Postman. Further, collaborated with team members on DTS AI Studio, used APM namespace in Kubeflow for time series analysis. Implemented machine learning techniques such as PCA, KNN, and Autoencoders for anomaly detection.

6.4.1) Assumptions

- Assumed access to DTS AI Studio, Kubeflow, and relevant datasets.
- Assumed familiarity with machine learning techniques like PCA, KNN, Autoencoders and anomaly detection.
- Access to Iris Foundry platform and Workbench platform.

6.4.2) Design Analysis and Experimental Work

➢ Objective

- Finalize P&ID tasks
- Explore FMEA templates
- Understand the centrifuge data problem
- Implement machine learning techniques for anomaly detection.

➢ Approach

- Cleaned and preprocessed data for analysis.
- Performed resampling, removed duplicates and dropped NaNs.
- Created channels and units on Postman for data management and matched the data.
- Used APM namespace in Kubeflow for time series analysis.
- Collaborated with team members on DTS AI Studio for centrifuge data analysis.
- Applied PCA, KNN, and Autoencoders for detecting anomalies in the data.

6.4.3) Tools and Libraries Used

- Python: scikit-learn for PCA and KNN, TensorFlow for Autoencoders
- Postman, Kubeflow, DTS AI Studio, Workbench

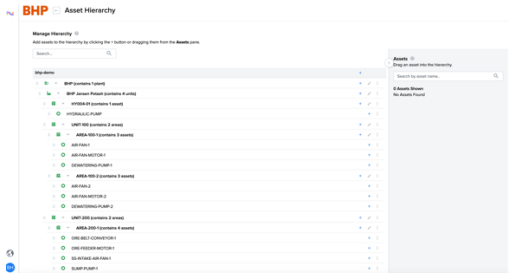

Fig. 30- BHP Asset Hierarchy

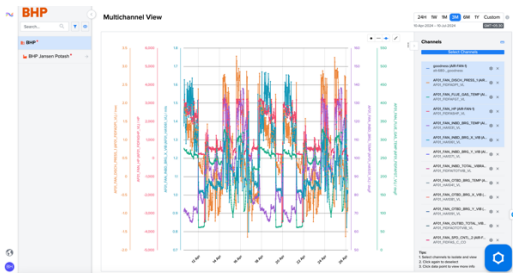

Fig. 31- BHP channel analytics

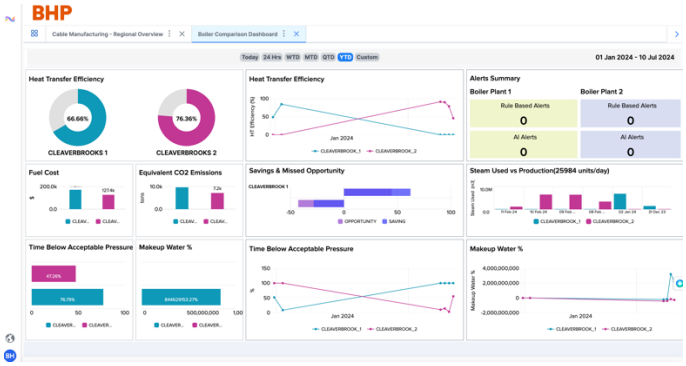

Fig. 32- Boiler analytics

6.4.4) Preliminary Results

- Cleaned and organized datasets for analysis.
- Successfully created channels and units for data management in Postman.
- Implemented PCA, KNN, and Autoencoders, identifying anomalies in the data.

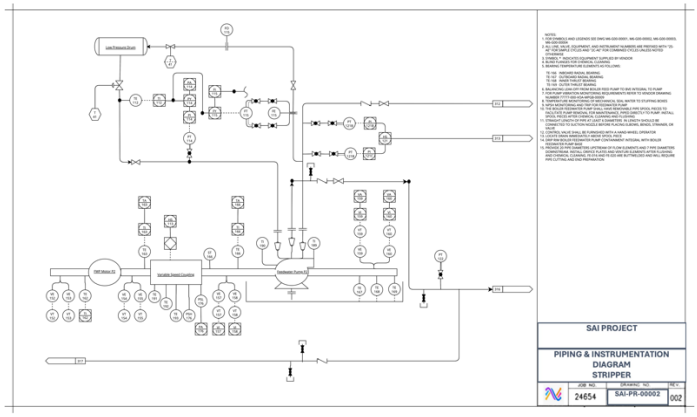

Fig. 33-Final P&ID

6.4.5) Eventual Results and Conclusions

- Finalized P&ID tasks with improved accuracy and efficiency.
- Generated valuable insights through advanced analysis techniques like time series visualization.
- Demonstrated the ability to apply machine learning for industrial process optimization.

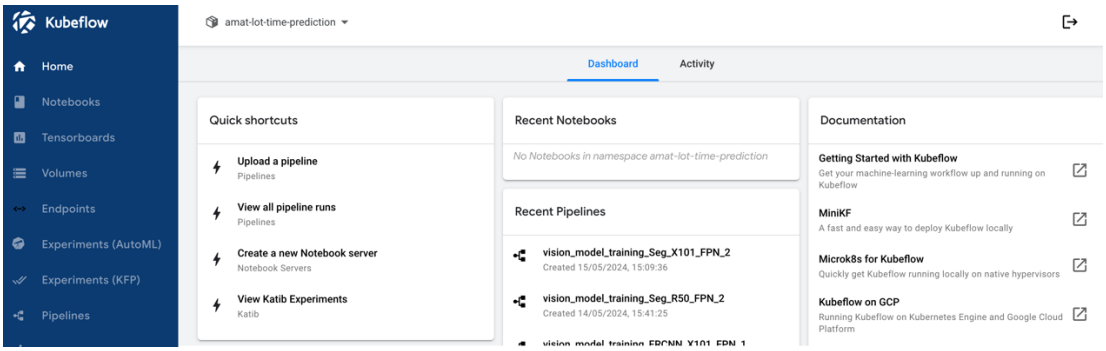

Fig. 34- Kubeflow portal

6.5) Working with the TDK and CF client problems

Worked on the individual machine data for the 5 machines. Pre-processed and cleaned data by removing the outliers. Further, visualised the data across boxplots and histograms to name a few techniques. The aim focused on comparing the variables across different datasets and provide statistical insights. The other problem created by the centrifugal data arise the need for anomaly detection techniques to improvise the cost and reduce wastes in machinery.

6.5.1) Assumptions

- Assumed access to Kubeflow, Workbench and relevant datasets.
- Assumed familiarity data cleaning and anomaly detection.
- Access to Iris Foundry platform and Workbench platform.

6.5.2) Design Analysis and Experimental Work

➢ Objective

- Understand the machine data problem.
- Work on the centrifuge data problem.
- Utilize all the platforms.

➢ Approach

- Performed general data pre-processing like resampling, removed duplicates and dropped NaNs.
- Outlier Removal using IQR (Inter-Quartile Range) to identify and remove outliers. Generally, outliers are defined as data points that lie outside 1.5 times the IQR below the 1st quartile (Q1) and above the 3rd quartile (Q3).
- Focused on data storage using a dictionary, allowing the use of only cleaned data.
- Visualization techniques generated boxplots and histograms using python libraries.
- Implemented KNN imputation to fill in missing values using neighbours 5 and 10.

6.5.3) Tools and Libraries Used

- Python: scikit-learn for KNN imputation, Seaborn and matplotlib for visualization plots, numpy, pandas, zscore
- Postman, Kubeflow, DTS AI Studio, Workbench

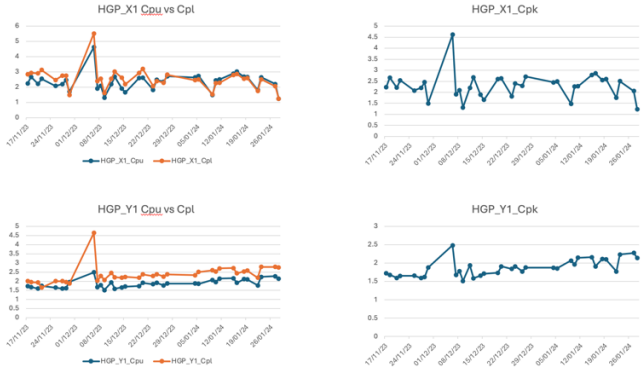

Fig. 35- Variable data for every machine

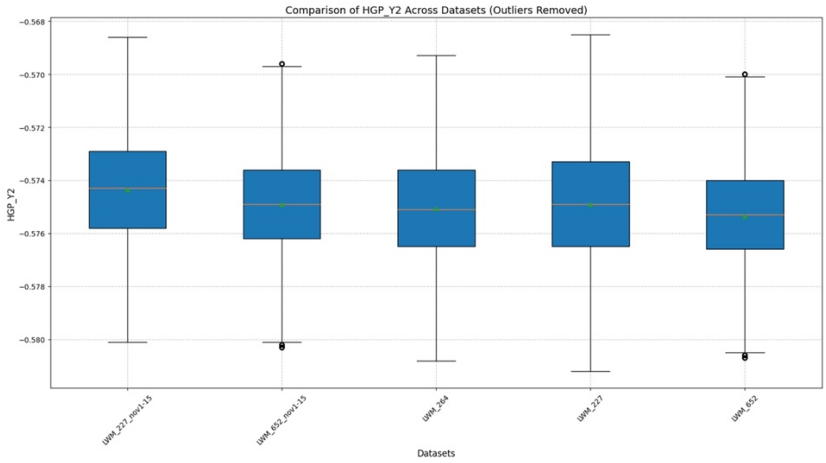

Fig. 36- Boxplots for each variable data on every machine

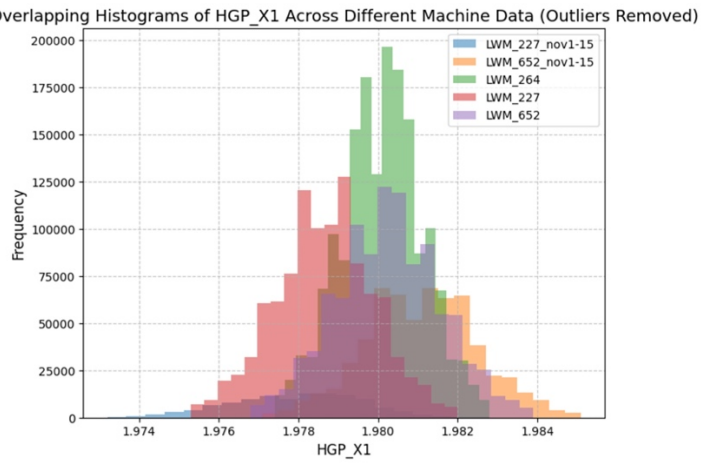

Fig. 37- Histograms for each variable data on every machine

6.5.4) Preliminary Results

- Successfully identified and removed outliers using the IQR method, ensuring cleaner data for analysis.
- Generated and displayed boxplots and histograms for each variable, enabling a clear visual comparison across different datasets.
- Implemented data cleaning, visualization, and anomaly detection.

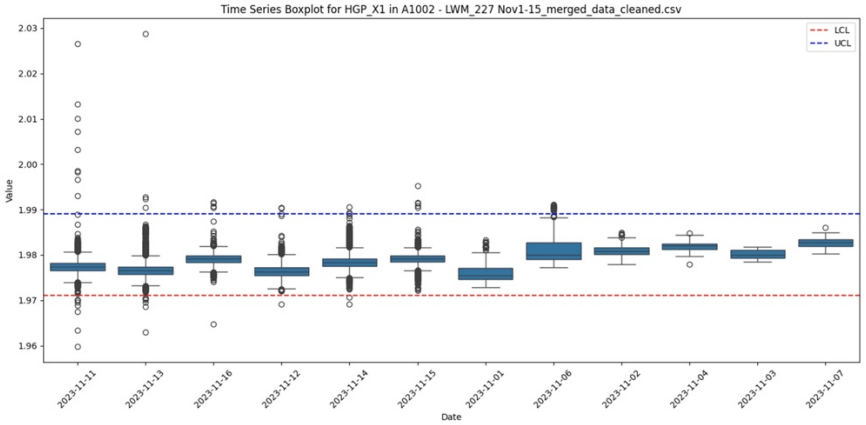

Fig. 38- Time series analysis for each variable data on every machine

6.5.5) Eventual Results and Conclusions

- Presented the data with improved understanding to the client problem.
- Generated valuable insights through advanced analysis techniques like time series visualization.
- Identified and removed outliers using IQR (Inter-Quartile Range) method, ensuring cleaner data for analysis.
- Generated and displayed boxplots and histograms for each variable, enabling a clear visual comparison across different datasets.
- This ensures filling missing data in the specified columns are accurately imputed and stored in a complex dataset for further analysis.
- Performed anomaly detection using the Workbench platform ADE (Anomaly Detection Engine).

6.6) Working with the UR and Nuvoko client issues

Worked with the UR problem and evaluated the problem points by a visualized analysis after performing data pre-processing. Moreover, targeted specific variables requiring customer issues.

6.6.1) Assumptions

- Assumed access to Kubeflow, Workbench and relevant datasets.
- Assumed familiarity data cleaning and anomaly detection.
- Access to Iris Foundry platform and Workbench platform.
- Assumed data cleaning processes and IQR statistical methods.
- Knowledge about EDA methods.

6.6.2) Design Analysis and Experimental Work

➢ Objective

- Understand the problem cases and utilize all the platforms.
- Read and pre-process time series data.
- Bring about visualized results after removing outliers in data.

➢ Approach

- Cleaned and preprocessed data for analysis.
- Filled merged cells by forward-filling merged cells filled NaN values and cleaned timestamp format to datetime.
- Calculate key statistics and control limits to generate scatter plots of the variable over time.
- Further, use the z-score method to identify and remove outliers from the data and calculate the improved control limits. The simplicity and ease of implementation, as well as the standardization comparison techniques preferred this method.

6.6.3) Tools and Libraries Used

- Python: Seaborn and matplotlib for analysis.
- Numpy, pandas, zscore

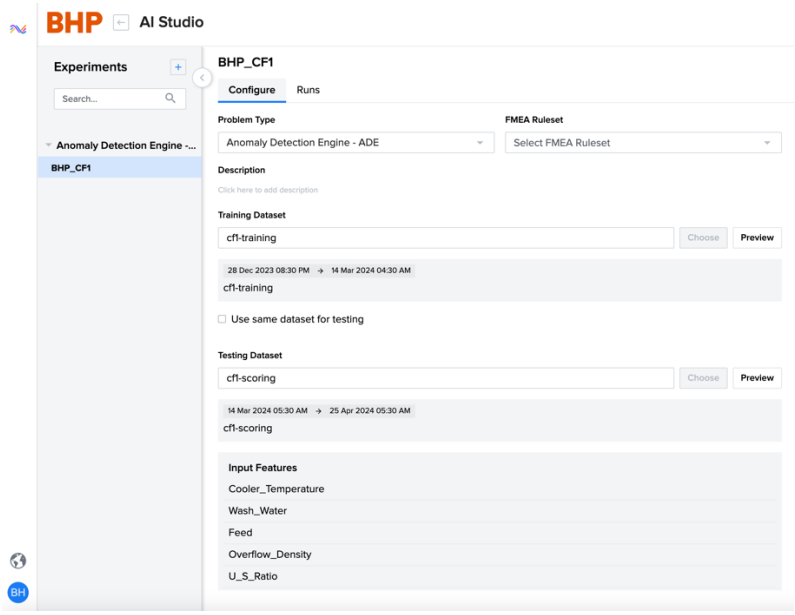

Fig. 39- CF data ADE

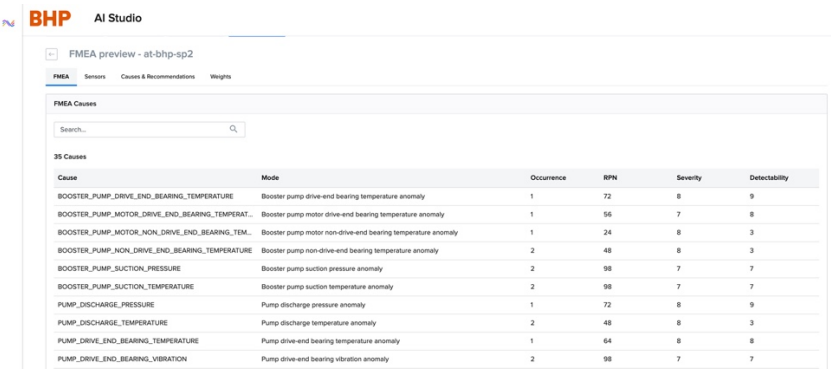

Fig. 40- FMEA platform

6.6.4) Preliminary Results

- Successfully identified and removed outliers using the IQR method, ensuring cleaner data for analysis.
- Generated and displayed boxplots and histograms for each variable, enabling a clear visual comparison across different datasets.
- Implemented data cleaning, visualization, and anomaly detection.

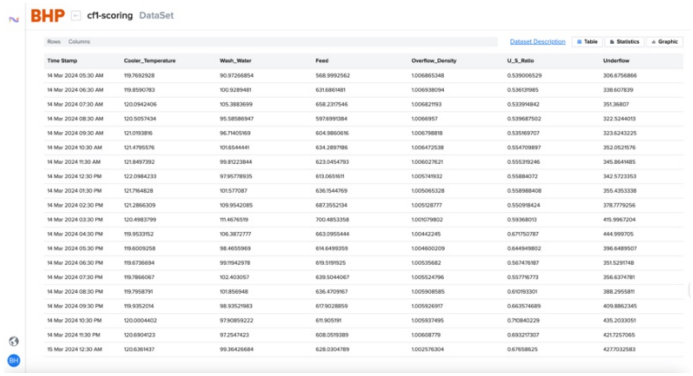

Fig. 41- scoring and testing data platform

6.6.5) Eventual Results and Conclusions

- Presented the data with improved understanding to the client problem.
- Generated valuable insights through advanced analysis techniques like time series visualization.
- Identified and removed outliers using IQR (Inter-Quartile Range) method, ensuring cleaner data for analysis.
- Generated and displayed boxplots and histograms for each variable, enabling a clear visual comparison across different datasets.
- This ensures filling missing data in the specified columns are accurately imputed and stored in a complex dataset for further analysis.

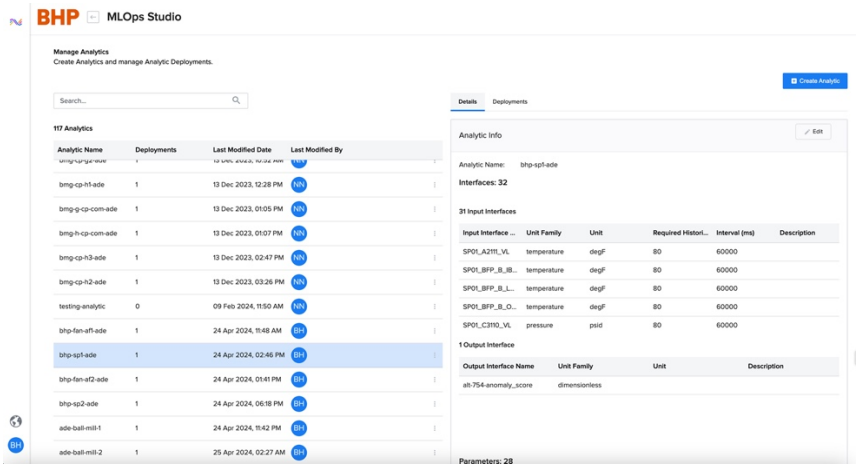

Fig. 42- Analytics platform for CF

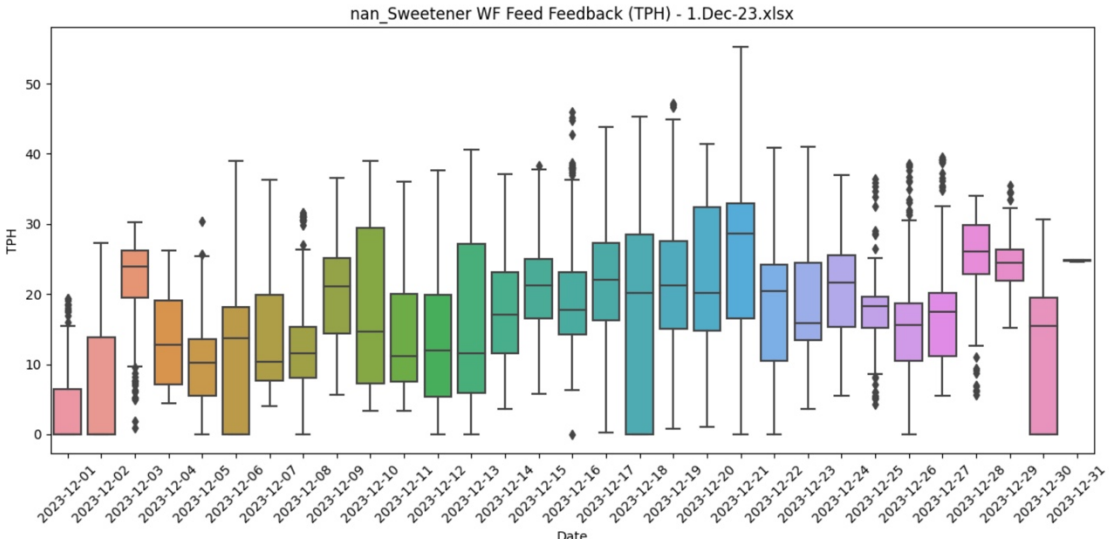

Fig. 43- Time series analysis for Nuvoko data

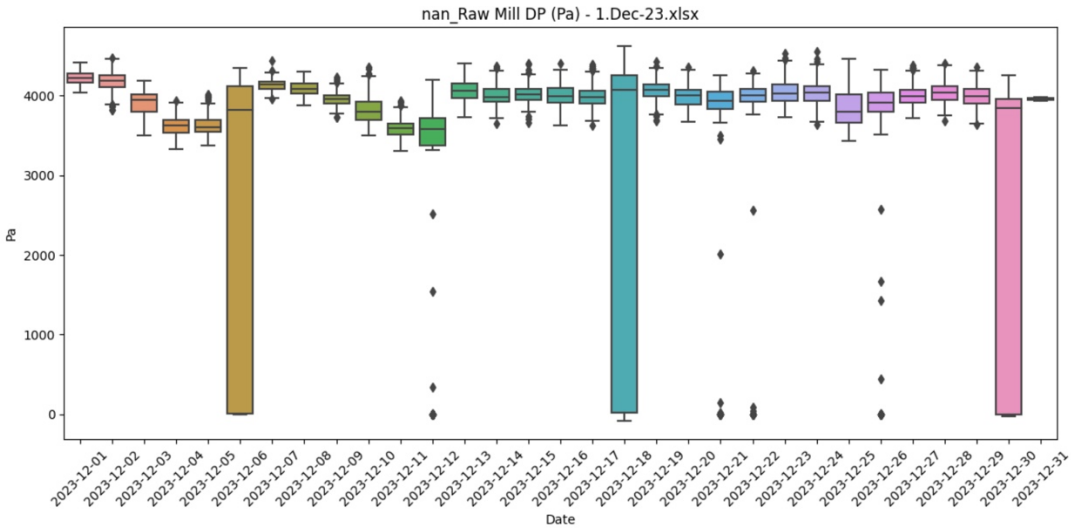

Fig. 44- Anomaly detection on Nuvoko data using time-series analysis

CHAPTER 7

RESULTS AND CONCLUSION

The statistical analysis and EDA performed on the data brought forward notable distinctions.

- **Foundation and Progression**
  Continuous improvement and learning were ensured by a detailed structured format of work. After initializing the work, each new phase and new task involved building on the previous task. Starting from basic understanding and setup, to finally progressing to advanced model application and detailed data analysis.

- **Technical Proficiency**
  The work undertaken enhanced technical proficiency in various tools and platforms. The SymphonyAI platforms included IRIS Foundry, Kubeflow, Energy and Petrochemical tenants Workbench, and DTS AI Predictive Portal. Further work involved understanding platforms and tools like Docker for containerization, Postman for API's, YOLO models for Computer Vision tasks like P&ID, and learning about Huggingface Transformers. Complex AI projects were handled with enhanced proficiency after learning and implementing these tools.

- **Data Quality and Analysis**
  Analysis consisted of implementing statistical analysis and EDA to understand the data as a visualized and structured solution. The meticulous correction of data and evaluating the ensured generating meaningful insights susceptible to the individual use cases.

- **Collaboration and Systematic Approach**
  The importance of teamwork collaboration and structured methodologies helped in understanding the bottleneck issues, and pragmatically implementing ADE to find anomalies in data.

CHAPTER 8

FUTURE SCOPE

Aim is to utilize the platforms developed by SymphonyAI to harness the use of AI in industrial optimization. This can be in automating error handling for P&ID diagrams to prevent any humanized delay in generating the diagram.

Further, GenAI (Generative AI) and PredAI (Predictive AI) can revolutionize the industry by enhancing operational efficiency and drive in innovation in the industry.

- Advanced predictive maintenance can enhance failure prediction and reduce wastage by real-time monitoring (predictive energy optimization).
- GenAI can optimize product generation and procedure.
- AI-assisted decision-making can help generating insights of complex industrial processes.
- Adaptive learning systems can gradually take over routine tasks by learning from human operators.

8.1) ENVIRONMENTAL AND SUSTAINABILITY PROCEDURES

- Perform anomaly detection to ensure anomaly detection is based on data collection issue and not on machinery failure.

8.2) SAFETY AND SECURITY PROCEDURES

- Real-time monitoring and anomaly detection further contributed to reducing resource wastage and power by understanding and stopping the machinery in case of any immediate problem.
- Safety and Security to maintain the confidentiality of data, and do not replicate the data inside or outside the work premises for any personal or financial gain.
- Intellectual Property: Ensured all designs, software, and innovations developed during the project complied with SymphonyAI's policies and respected third-party IP rights.
- Ethical Dilemma of comparing client data of the same problem to come at a faster solution versus remaining confidentiality of individual data.

REFERENCES

[1] Kimray Inc., [Online] Available: https://www.youtube.com/c/KimrayInc on April, 2024

[2] Kimray Inc., P&ID reference guide, Available: Kimray on April, 2024

[3] Data Cleaning, [Online] Available: geeksforgeeks

[4] Learning postman, Docs overview, https://learning.postman.com/docs/introduction/overview/ in March 2024

[5] Anomaly Detection, IBM [Online] Available: https://developer.ibm.com/learningpaths/ in May 2024

Figures

[1] Plant Insights
[2] AI powered smart manufacturing
[3] IRIS Foundry
[4] IRIS Vision AI
[5] EurekaAI platform
[6] CNN architecture
[7] FMEA analysis
[8] FMEA chart
[9] KNN distance metrics
[10] Autoencoder architecture

Annexure 1
PO &PSO Mapping

Student Name: Aryaman Singh Dev
Registration no: 200929064
Note: use a tick mark if you have addressed that PO in your report

PO	· Tick	Pg. No	Section No	Guides Observation
PO1	√		CHAPTER 5 AND 6	
PO2	√		CHAPTER 5 AND 2	
PO3	√		CHAPTER 6	
PO4				
PO5	√		CHAPTER 5, 6 AND 7	
PO6				
PO7	√		CHAPTER 1 AND 7	
PO8	√		CHAPTER 1, 5 AND 7	
PO9	√		CHAPTER 6	
PO10				
PO11	√		CHAPTER 6	
PO12	√		CHAPTER 8	

PSO	· Tick	Pg. No	Section No	Guides Observation
PSO1				
PSO2				
PSO3	√		CHAPTER 6	

Signature of Student: 

Name and Signature of Guide: Mr. Mahesh Anil Inamdar

Date: 13 August 2024

Annexure 2
PLO Mapping

Student Name: Aryaman Singh Dev
Register no: 200929064
Note: use a tick mark if you have addressed that LO in your report

Sl	PLO	· Tick	Pg. No	Section No	Guides Observation
1	C1.	√	1-12, 21-40	CHAPTER 1,2 AND 6	
2	C2.	√	1-12, 21-40	CHAPTER 1,2 AND 6	
3	C3.	√	17-40	CHAPTER 5 AND 6	
4	C4.	√	15-40	CHAPTER 4,5,6	
5	C5.	√	17-40, 42-43	CHAPTER 5, 6 AND 8	
6	C6.				
7	C7.	√	42-43	CHAPTER 8	
8	C8.				
9	C9.				
10	C10.	√	42-43	CHAPTER 8	
11	C11.				
12	C12.	√			
13	C13.	√			
14	C14.	√	17-40	CHAPTER 5 AND 6	
15	C15.	√			
16	C16.	√	17-40	CHAPTER 5 AND 6	
17	C17.	√	21-40, 42-43	CHAPTER 6 AND 8	

18	**C18.**	√			

Signature of Student:

Name and Signature of Guide:

Mr. Mahesh Anil Inamdar

Date: 13 August 2024

Annexure 3

Address of IET learning outcomes during project period

Answer the following questions relevant to your Practice School work.

1. Explain the steps you followed to Investigate and define the problem in your project work (C4, evaluate level)

Answer: To investigate and define the problem, initial meetings were held to understand existing challenges and objectives required to be fulfilled for the clients. A thorough review of P&ID diagrams, data collection and cleaning, machine learning algorithms, anomaly detection systems and data visualization techniques. Root cause analysis revealed the need for automated control and real-time monitoring which was fulfilled by time-series analysis of the data. A clear problem statement was formulated, objectives set, and risk assessments conducted.

2. What is the science, mathematics, statistics, engineering principles and other basic technology you identified for design (Mechanical, Electronic, Physics, Chemistry, Automation) in your project work? (C1, C2, C3, Application, Analysis, Evaluation of Science and Mathematics in the project)

Answer: Science involved understanding electrical and mechanical systems. Mathematics was used for calculating the statistics and calculations of data. Engineering principles included mechanical engineering for designing the P&ID diagrams, electronics to understand the problem cases. Computer Science skills were required to perform analysis of the data and find the root causes of the problem.

3. Have you considered the Environmental and sustainability limitations in your project work? (C7, evaluate)

Answer: Yes, environmental and sustainability limitations were considered in the project. The P&ID diagrams generated ensured adhering to the principles in real-time. Real-time monitoring and fault detection further contributed to reducing resource wastage and power by understanding and stopping the machinery in case of any immediate problem.

4. Have you considered ethical dilemma, health, safety, security, and risk issues; intellectual property; codes of practice and standards? Did you address any of these issues in your project work? If so, Explain in detail. (C5, create)

Answer: Safety and Security: Remain the confidentiality of data, and do not replicate the data inside or outside the work premises for any personal or financial gain.

Intellectual Property: Ensured all designs, software, and innovations developed during the project complied with SymphonyAI's policies and respected third-party IP rights.

Ethical Dilemma: Comparing client data of the same problem to come at a faster solution versus remaining confidentiality of individual data.

5. What were the esthetical issues faced and how it is addressed in your project in the design phase? (C5, analysis)

Answer: Issues primarily focused on integrating new techniques to accelerate the growth of the company. Challenges included maintenance issues and real-time analysis for identifying the root cause, understanding the working of P&ID systems. To address these issues, the following was completed:

P&ID – Focusing on building efficient solutions, that can lower cost, wastage and improve optimization.

Real-Time Analysis – Analysing previous trends to predict future outcomes and ensure machinery is up to date.

6. Were there any health issues considered during the design process. How is it addressed in your project in the design phase? (C5, create)

Answer: My work was purely software, so no health issues during the design phase.

7. What were the safety, security and risk issues needed to be taken care of in the design stage? (C10, create)

Answer: Security of data was required for remaining confidentiality of clients. This was completed by a dual factor login platform, enhancing the security of uploaded data.

Risks associated were client data leakage, which was prevented by utilizing the enhanced security of the SymphonyAI platforms.

8. Were there any intellectual property issues that needed to be taken care off? Have you come across IP issues in the project phase? (C5, create)

Answer: Intellectual property is taken care by citing and proper referencing. So, no issues faced.

9. What are the codes of conduct and standards you needed to use in the design phase and in other phases of your project as well? (It may include codes of practice and standards for safety, security, health, risk) Explain the legal issues, ISO standards, IEC standards, etc. (C8, evaluate)

Answer: During my internship as a Data Scientist, I had the opportunity to work on several phases of data handling, from cleaning to visualization. Throughout these processes, I adhered to various codes of conduct and standards to ensure safety, security, and compliance with legal requirements. Here are some key standards and practices I followed:

i. Data Privacy and Confidentiality

- GDPR (General Data Protection Regulation):

This regulation is critical for handling any data originating from the European Union. It ensures the protection of personal data and privacy. I ensured that any personal data was anonymized or pseudonymized and that data processing activities were transparent to the data subjects.

- HIPAA (Health Insurance Portability and Accountability Act):

Relevant for projects involving health data, HIPAA sets standards for protecting sensitive patient information. While handling any health-related data, I followed the protocols to ensure data encryption and restricted access to authorized personnel only.

ii. Industry Standards

- ISO/IEC 27001 (Information Security Management):

This standard outlines the requirements for establishing, implementing, maintaining, and continually improving an information security management system (ISMS). I adhered to the ISMS policies to protect the confidentiality, integrity, and availability of data, ensuring robust security measures were in place.

- ISO 9001 (Quality Management Systems):

This standard focuses on meeting customer and regulatory requirements and enhancing customer satisfaction through effective quality management systems. I applied the principles of ISO 9001 to ensure high-quality data processing and analysis, consistently reviewing and improving our processes.

iii. Data Handling and Analysis

- ISO/IEC 2382 (Information Technology Vocabulary):

This standard provides a comprehensive vocabulary for information technology, ensuring clear communication and understanding across the industry. I used standardized terminology in documentation and reporting to maintain clarity and consistency in communication.

- ISO 26000 (Social Responsibility):

This standard provides guidance on how businesses and organizations can operate in a socially responsible way. I followed guidelines on ethical data handling, ensuring that our data practices did not adversely impact stakeholders or society at large.

iv. Legal Issues

- Intellectual Property Rights:

Ensuring that data sources, algorithms, and visualizations respect intellectual property laws. I avoided using proprietary data or software without proper licensing and gave appropriate credit to original sources.

v. Platforms and Tools

- Workbench for Anomaly Detection:

Workbench is used for detecting anomalies in data before further analysis. I followed best practices in anomaly detection, ensuring that the data was clean and reliable for subsequent analysis.

- Kubeflow for Data Visualization:

Kubeflow is a platform for deploying machine learning workflows, including data visualization. I maintained data confidentiality and security by ensuring that all visualizations were generated and shared in a secure environment, with access controls in place.

This ensured that the data handling processes during my internship were not only effective but also compliant with legal and ethical standards. This approach not only safeguarded the data but also built trust with clients and stakeholders.

10. What is the general safety measure regulated in the industry where you did the project work? (C8, evaluate)

Answer: In the industry where I completed my project, general safety measures were heavily regulated to ensure a safe working environment for all employees.

Workplace Safety Protocols included regular safety drills and training sessions for employees to handle emergencies such as fire, electrical hazards, and data breaches. Data Security Measures consisted of implementation of secure access controls to protect sensitive data from unauthorized access. This allowed regular updates and patches for software to prevent security vulnerabilities.

11. What were the professional ethics needed to be followed in general while you are doing the project? (C8, evaluate)

Answer: While working on the project, it was essential to follow certain professional ethics, which included integrity, confidentiality, and accountability and respect.

- Integrity ensures transparency in all data handling and reporting processes, thus avoiding any form of data manipulation or misrepresentation.
- Moreover, it is necessary to maintain the confidentiality of client data and respecting privacy agreements, thereby ensuring that sensitive information was only accessible to authorized personnel.
- Taking responsibility for the quality and accuracy of the work actively builds accountability.
- Respecting colleagues, clients, their contributions and valuing diversity and promoting an inclusive work environment.

12. Do you think ethics and professionalism needs to be paid attention to by students during study? If yes, explain how it can be inculcated/introduced/implemented? (C8, evaluate)

Answer: Yes. Students should be given a lecture on ethics and professionalism or asked to do some online courses for the same. As this is not taught much in our schoolings, through case studies also this topic and its importance could be explained to the students.

- Curriculum Integration ensures that students gain a thorough understanding of professional ethics by including relevant courses such as corporate social responsibility and legal issues. Case studies and scenarios should be provided to help students grasp the practical implications of ethical conduct.

- Workshops and Seminars organized with industry professionals can discuss the importance of ethics and professionalism. Students are encouraged to participate in ethics competitions and debates, deepening their understanding.

- Internships and Real-world Projects offer students opportunities to emphasize ethical considerations and professional behaviour. Mentorship programs allow students to learn from experienced professionals, fostering a commitment to ethics in their future careers.

13. Do you think environmental and sustainability limitations; ethical, health, safety, security, and risk issues; intellectual property; codes of practice and standards are sufficiently covered in the courses you have studied in your curriculum? (C8, evaluate)

Answer: Each course related to industry in the curriculum includes at least one or two sessions focused on ethics and/or environmental sustainability.

- Case Studies can enhance the curriculum by incorporating more real-world examples that highlight the importance of environmental and sustainability limitations, as well as ethical, health, safety, security, and risk issues.

- An Interdisciplinary Approach can be promoted by combining technical education with courses in ethics, law, and environmental science, providing a more holistic understanding of these critical topics.

- Practical Applications should be encouraged through projects and hands-on experiences that directly address these limitations and issues, ensuring that students can apply their knowledge in real-world scenarios.

14. Have you gone through online classes, or a crash course in which you are familiarized with intellectual property rights as well as risk issues in a professional environment? (C8, evaluate)

Answer: Yes, some work did include these aspects.

15. In the beginning of your project did you evaluate environmental effects and sustainability factors in your work? (C7, evaluate)

Answer: No, I was not informed in the beginning

16. During your stay in the industry, have you realized the need for professional and ethical conduct? Quote the context and explain. (C8, evaluate)

Answer: Yes, the need for professional and ethical conduct is a must in any industry for growth. Work cannot be conducted in any industry with discipline without professional and ethical conduct.

During my stay in the industry, I realized the importance of professional and ethical conduct. For example, while handling sensitive client data, I had to ensure strict confidentiality and follow all data protection regulations. This experience underscored the need for ethical decision-making for decisions that prioritize ethical considerations over shortcuts or expedient solutions and compliance for adhering to all legal and regulatory requirements, even when under pressure to meet tight deadlines.

17. What are the professional codes of conduct you needed to imbibe during your stay in the industry? (C8, evaluate)

Answer: Integrity, confidentiality, competence, conflict of interest, respect for others.

Professional codes of conduct included,

- Confidentiality Agreements by signing and adhering to confidentiality agreements to protect client data.
- Maintaining professionalism in all interactions with clients and colleagues.
- Staying updated with the latest industry standards and best practices and following continuous learning approach

18. Did you address any of limitations of your project work and have you improved the results through continuous improvements in your project work? (C5, create)

Answer: I went through different research papers and wrote the research gap. This led to the objectives of the project. Biweekly and monthly reports were made to track my progress.

19. How did you plan your project, deadlines, maintaining dairy of each stage and improved the quality of the project (C14, understand)

Answer: Breaking down of tasks into small target with defined deadlines, continuous discussions with the supervisor for progress or places where change of approach is required.

To plan my project and improve its quality, I followed these steps like regular project planning to set clear goals and deadlines for each stage of the project. Moreover, keeping a detailed diary of each stage of the project to document challenges and solutions allowed flexibility.

20. While having Industrial training/ internship, what were the college practices which helped you to abide by the professional ethics of the company environment? (C8, evaluate)

Answer: Adhering to the deadline for submission was one of the main college practices which helped in abiding by the professional ethics of the company environment.

Several college practices helped me abide by the professional ethics of the company environment, such as adhering to the college's code of conduct, which emphasized integrity and ethical behaviour, receiving guidance from faculty mentors on professional and ethical conduct, and participating in industry interaction sessions that highlighted the importance of ethics in the professional world.

21. During your stay in the industry, did you observe how teamwork plays a role in engineering process? (C16, apply)

Answer: Yes, working in a team provides you with support and you also gain knowledge when part of a healthy discussion.

During my stay in the industry, I observed how teamwork plays a crucial role in the engineering process. For instance:

- Collaborative Problem-Solving: allowed for diverse perspectives and collaborative problem-solving, leading to more innovative solutions.
- Division of Labor: ensured each team member contributing their expertise to the project.
- Communication and tracking: facilitated coordination and alignment of goals within the team.

22. Are you aware of the ethical clearance when you work in the field of health/medical applications? (C8, evaluate)

Answer: Yes, it involves getting approval from the ethics committee to ensure research project meets ethical standards and human well-being. This involves data collection, consent forms and relevant documentation.

23. During your stay in industry, are you able to observe and understand certain management techniques practiced in that industry. Explain in detail. (C14, understand)

Answer: Yes, some of the management techniques were focusing on continuous improvement of product, quality management of products, benchmarking of the similar solutions offered by competitors. Along with this, lean management and the agile methodology offered insights on performance metrics to track progress and performance.

24. Could you understand how they tackle project management and what tools and techniques are adopted? (C14, understand)

Answer: SymphonyAI employed several tools and techniques for effective project management.

- Project Management Software for task management, sprint planning, and tracking project progress.
- Risk Management Techniques to identify, evaluate, and mitigate potential risks throughout the project lifecycle.

25. During your stay in the industry, did you observe any engineering activity implemented to promote sustainable development? (C7, evaluate)

Answer: No observations on this.

26. Did you adopt any quantitative technique for any engineering activity related to your project? (C3, evaluate)

Answer: Yes, I adopted quantitative techniques for engineering activities related to my project.

- Statistical Analysis using techniques like regression analysis and hypothesis testing were used to identify trends and relationships in the data. Z-score calculations were employed for outlier detection in data preprocessing.
- Supervised Machine Learning Algorithms such as linear regression and decision trees, were used to predict outcomes and derive insights from the data.

27. What are the elements of your project work which addresses sustainable development and were you able to apply quantitative techniques to analyze and achieve your project goals? (C7, evaluate)

Answer: Elements of my project that addressed sustainable development included:

- Energy Efficiency to optimize data processing algorithms to reduce computational resource consumption.
- Quantitative Techniques to analyze data and derive insights that could lead to more sustainable practices in data management.

28. How the company takes green initiative, environment related factors. (C7, evaluate)

Answer: Using anomaly detection allows us to find points of failure by utilizing predictive AI techniques, thereby affecting recycling of waste and minimization of the waste generated.

29. During your stay in the industry, have you observed/sensitized about legal requirements governing such activities in that industry? Explain. (C8, evaluate)

Answer: No

30. Did your project need the understanding of relevant legal requirements governing engineering activities you carried out as a part of your project work? Explain in detail. (C8, evaluate)

Answer: No

31. What are the legal and ethical practices you followed while working on the project? (C8, evaluate)

Answer: No such practice was carried out.

32. Are you sure that you abide IPR/copy right issues? (C15, apply)

Answer: Yes, Intellectual property is taken care by citing and proper referencing. So, no issues will arise.

33. Have you observed any national/international standards in the workplace? How many are relevant to your project work? List them. (C8, evaluate)

Answer: Relevant standards to my project work included:

- ISO/IEC 27001: For information security management.
- GDPR: For data protection and privacy.

34. What online course you attended to improve your communication skills. Report writing, Oral presentation, Software used for writing report. (C17, apply)

Answer: No course was taken. Everything in the report was done by prior experience in report writing.

35. In your project, was it needed to tackle risk issues, including health & safety, environmental and commercial risk, and risk assessment and risk management techniques? Explain in detail. (C5, create)

Answer: No, there was no such need.

36. What are the cyber safety rules and precautions you were sensitized with, when you started practice school, or started industrial training? (C9, evaluate)

Answer: Few cyber safety rules were to not send any documents to an outside source not verified by the company and not to upload any company content online.

37. How is an organization addressing a fire accident/human safety when working with machines? (C9, evaluate)

Answer: Fire extinguishers at entrance of office and in halls inside with instructions on how to use them. Precautions to be taken during fire are also cleared before joining.

38. Process of teamwork. How each of you are involved in the team? What part of the work is addressed by you.? (C16, evaluate)

Answer: I was primarily responsible for data cleaning and anomaly detection using Workbench, as well as data visualization using Kubeflow. Team members collaborated on tasks through regular meetings and shared documentation.

39. Have you filed a patent, IPR, or published your work? Give more details. (C17, evaluate)

Answer: No patent has been filed yet.

40. How you documented the literature review, your analysis on their results, discussion with the guide and team members, provide the documents on a weekly basis. Put as one chapter in final report. (C4, evaluate).

Answer: This has been included in the final report as chapter 2. All the above-mentioned analysis has been done in that chapter.

41. Have you sensitized about inclusion and diversity in the team? If yes, what are the diversification in the team in terms of religion, gender, ethnicity, etc. What challenges you come across in the team. (C11, apply). Indian constitution and acts related to caste, gender, race discrimination.

Answer: Our team was already diverse in terms of gender, ethnicity, and cultural backgrounds.

42. How were you able to keep yourself updated with the technology? How you incorporated advanced technology in your project. (C18, lifelong learning)

Answer: A variety of technologies were utilised to stay in touch with recent advancements.

- Taking courses on platforms like Coursera to learn about new tools and technologies.
- Attending webinars and industry conferences to stay informed about the latest trends.
- Implementing machine learning algorithms and data visualization tools in my project.

43. Which is the laboratory skill you found applicable to your project. Explain.

(C12, apply)

Answer: Skills like data collection, data pre-processing, anomaly detection, documentation and software utilization were useful in the project.

Annexure 4

Project/Practice School Classification

Student Name: Aryaman Singh Dev
Registration no: 200929064

Note: Use a tick mark to specify under which domain your practice schoolwork falls into.

Table 1: classification based on project domain classification

Domain	Tick
Product	√
Application	√
Review	
Research	
Management	

Note: Use a tick mark to specify Societal impacts you considered during your practice school.

Table 2: classification based on societal consideration

Societal Impact	Tick
ethics	√
safety	√
environmental	√
commercial	√
economic	√
social	

Signature of Student:

Name and Signature of Guide:

Mr. Mahesh Anil Inamdar

Date: 13 August 2024

Annexure 5

Company Details

Student Details			
Student Name	Aryaman Singh Dev		
Register Number	200929064	Section/Roll No	A/22
Email Address	Aryaman.dev@learner.manipal.edu	Phone No (M)	8851759219
Project Details			
Project Title	GENERATIVE AND PREDICTIVE AI FOR IMPROVED INDUSTRIAL PROCESSES		
Project Duration	5.5 Months	Start Date	15TH Jan 2024
End Date	30TH June 2024		
Organization Details			
Organization Name	SymphonyAI		
Type of Organization (Public Listed, Private, PSU, Govt, cooperative)	Private		

Full postal address with pin code	1, EPIP Zone, Whitefield, Bengaluru, 560037		
Website address	https://www.symphonyai.com/		
Name of the CEO of the Organization	Sanjay Dhawan		
Supervisor Details			
Supervisor Name	Sudhanshu Kashyap		
Designation	Director - AI		
Full contact address with pin code	Brigade Golden Triangle Bangalore		
Email address	sudhanshu.kashyap@symphonyai.com	Phone No (M)	+91 9663122662
Internal Guide Details			
Faculty Name	Mr. Mahesh Anil Inamdar		
Full contact address with pin code	MIT Manipal, Udupi - Karkala Rd, Eshwar Nagar, Manipal, Karnataka 576104		
Email address	Mahesh.inamdar@manipal.edu		

3% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

- **13** Not Cited or Quoted 3%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **2** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
1% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **13** Not Cited or Quoted 3%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **2** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
1% Publications
1% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
www.epicos.com		1%
2	Internet	
www.eenewseurope.com		0%
3	Student papers	
University of Limerick		0%
4	Student papers	
Manipal University		0%
5	Internet	
docplayer.net		0%
6	Student papers	
Colorado Technical University Online		0%
7	Internet	
krishikosh.egranth.ac.in		0%
8	Internet	
fastercapital.com		0%
9	Internet	
journal.eu-jr.eu		0%
10	Publication	
H L Gururaj, M R Pooja, Francesco Flammini. "Recent Trends in Computational Sci...		0%

turnitin Page 3 of 52 - Integrity Overview Submission ID trn:oid:::1:2981079499

11 Internet
publications.polymtl.ca 0%

12 Internet
pure.manchester.ac.uk 0%

13 Internet
www.mdpi.com 0%

turnitin Page 4 of 52 - Integrity Overview Submission ID trn:oid:::1:2981079499