ABSTRACT

This report summarizes the work accomplished during my internship under the project titled "BSP Project on Pre-failure Alerting in BRM" at IIT Bhilai, supervised by Dr. Gagan Raj Gupta. The internship, initially scheduled from February 22, 2024, to May 22, 2024, was extended by three weeks, culminating on June 12, 2024.

The primary focus of the internship was the application of data analytics techniques, specifically motif and discord analysis, to enhance the reliability and efficiency of equipment at the BSP mill. Utilizing Python, I conducted motif analysis to identify normal patterns and discord analysis to detect anomalies within the equipment data. These analyses were crucial in preempting equipment failures and reducing cobbles, thereby contributing to the overall efficiency of the mill operations.

A significant part of my contributions involved augmenting the capabilities of the Grafana dashboard used for data visualization. I integrated the identified motifs into the dashboard to provide clear visual representations of normal operating patterns. Additionally, I implemented a dynamic discord detection feature that utilized real-time data from the BSP mill server to highlight potential anomalies. This feature was complemented by the creation of bar graphs to visualize binary signals, further enhancing the dashboard's functionality.

Moreover, I engaged in case studies that examined specific anomaly events identified from the BSP mill server data. These case studies included both true positive and false positive scenarios, providing comprehensive insights into the performance and accuracy of the anomaly detection system. The analysis of false positives was particularly valuable in refining the system to reduce erroneous alerts and improve its reliability.

Through these efforts, my internship significantly advanced the project's objectives, demonstrating the practical application of data analytics in industrial settings and contributing to the development of a robust pre-failure alert system. This experience has equipped me with valuable skills in data analysis, real-time data visualization, and the practical implementation of machine learning techniques in a real-world context.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1.	**DTW**	-	Dynamic Time Warping
2.	**RGB**	-	Red, Green & Blue
3.	**PNG**	-	Portable Network Graphics

CHAPTER – I

INTRODUCTION

CHAPTER – I

INTRODUCTION

1.1 Background Information

This report details the work accomplished during my internship on the project titled "Digital Transformation System for Pre-Failure Alert Generation for Equipment Failure & Cobble Reduction Based on Data Analytics and Video Analytics at BRM" at IIT Bhilai, supervised by Dr. Gagan Raj Gupta. The project focuses on utilizing advanced data analytics techniques to enhance the reliability and efficiency of equipment at the Bhilai Steel Plant's (BSP) Bar and Rod Mill (BRM). By analyzing patterns in equipment data, the project aims to preempt failures and reduce operational disruptions, thereby improving overall productivity. By analyzing patterns in equipment data, the project aims to preempt failures and reduce operational disruptions, thereby improving overall productivity.

1.2 Project Objectives

The main objectives of the project are:

- To conduct motif analysis to identify normal operating patterns within the equipment data.
- To perform discord analysis to detect anomalies that may indicate potential equipment failures.
- To enhance the Grafana dashboard for better visualization of data patterns and anomalies.
- To integrate motifs and implement dynamic discord detection within the dashboard.
- To create visual tools, such as bar graphs, to represent binary signals.
- To analyze both true positive and false positive case studies of anomaly events to refine the detection system.

1.3 Significance of the Project

This project is highly significant within the field of industrial data analytics. By developing a robust pre-failure alert system, it addresses a critical need for improving equipment reliability and operational efficiency. The ability to predict and prevent equipment failures can lead to significant cost savings and productivity gains in industrial settings. Furthermore, the integration of data analytics and video analytics represents a cutting-edge approach to industrial monitoring and maintenance.

1.4 Scope and Limitations

The scope of the project includes:

- Analyzing data from the BSP mill server to identify patterns and anomalies.
- Enhancing the Grafana dashboard to visualize these patterns and anomalies effectively.
- Conducting case studies to validate the anomaly detection system.
- The project is limited by:
- The irregularity of signal patterns, which can complicate the analysis.
- The presence of vibrations and spikes in the data, which can lead to false positives in anomaly detection.

1.5 Overview of the Structure

This report is organized as follows:

- **Introduction**: Provides background information, project objectives, significance, scope, limitations, and an overview of the report structure. The objectives are clearly stated, such as pre-empting failures and reducing cobble production through anomaly detection. The significance of the project is emphasized in terms of cost savings, improved production quality, and increased safety. Finally, the report structure is presented, providing a roadmap for the reader.

- **Data Decoding**: Details the methods used for analyzing and interpreting the equipment data. This raw data often requires cleaning before analysis.
- **Motifs and Discord Analysis**: Discusses the processes and results of identifying normal patterns and anomalies in the data and their integration.
- **Dashboard Enhancements**: Describes the improvements made to the Grafana dashboard for better data visualization.
- **Case Studies**: This section showcases the real-world effectiveness of the system through detailed analyses of specific events. It presents a case study of a "true positive" anomaly, where an alert led to the identification and rectification of a potential equipment failure. Conversely, a "false positive" case study explores an instance where an alert triggered an unnecessary investigation. Both scenarios highlight the system's performance and areas for improvement.
- **Conclusion**: Summarizes the findings, discusses the implications, and suggests future directions for the project. Finally, the section suggests potential future directions for further development and improvement. This could involve incorporating additional data sources, exploring more sophisticated anomaly detection techniques, or integrating machine learning.

CHAPTER – II

ORGANIZATION OVERVIEW

CHAPTER II

ORGANIZATION OVERVIEW

2.1 Institute Overview

- **History:** Established in 2016 by the Ministry of Education (India) as one of six new Indian Institutes of Technology (IITs).
- **Size:** Relatively new and growing, with a student body expected to reach 12,000 in the coming years.
- **Location:** Bhilai, Chhattisgarh, India.
- **Logo:** A blue hexagon with an interconnected network design, symbolizing innovation and technological synergy.

Fig 2.1: IIT Bhilai Logo

- **Industry:** Education and Research - Focused on Science, Technology, Engineering, and Mathematics (STEM) fields.
- **Reputation:** Gaining recognition as a premier institute for technological education and research in India.

2.2 Vision and Mission

- **Vision:** To play a significant role in national development through technological advancements.

- **Mission:**
  - To provide a stimulating environment for learning and research in technology and related fields.
  - To foster innovation and originality.
  - To nurture well-rounded graduates equipped to address national challenges.

2.3 Organizational Structure

- **Governance:** Managed by the Board of Governors (BOG) and the Senate. Director heads the institute, supported by Deans and Heads of Departments.
- **Departments:** Offers various undergraduate and postgraduate programs in departments including Electrical Engineering, Computer Science, Mechanical Engineering, Physics, Chemistry, Mathematics, Liberal Arts, and Mechatronics.

2.4 Products and Services

- **Educational Programs:** Provides BTech, MTech, MSc, and PhD programs in Engineering, Science, and Liberal Arts disciplines.
- **Research Activities:** Conducts research in various technological fields with a focus on real-world applications and industrial collaborations.
- **Knowledge Transfer:** Aims to bridge the gap between academia and industry through consultancy services and technology development projects.

CHAPTER – III

INTERNSHIP ACTIVITIES

CHAPTER III

INTERNSHIP ACTIVITIES

This section outlines the core activities undertaken during my internship, the skills developed, challenges encountered, achievements made, and key learnings acquired.

3.1. Description of Activities

- **Data Decoding:** I facilitated the transformation of mill data from a complex, proprietary format (.dat) into a readily usable format for analysis using IBA Analyzer software. This involved data extraction and transformation for specific days spanning January to April.
- **Motif and Discord Analysis:** My core responsibilities included the identification of recurring patterns ("motifs") and deviations ("discords") within the time series data. I employed a multi-pronged approach, utilizing z-normalized Euclidean distance and matrix profile calculations for motif identification. Discrepancy detection leveraged Euclidean distance combined with Dynamic Time Warping (DTW) for its superior handling of time series variations. Establishing optimal thresholds streamlined visualization and effectively differentiated anomalies from motifs. Additionally, I optimized the window size for motif identification through theoretical analysis and practical case studies involving true and false positives.
- **Dashboard Enhancements:** My contributions extended to enhancing the data visualization dashboards. Binary signals were readily identifiable based on their period (".") notation. I implemented bar charts to effectively showcase the durations of these signals for clear comparison. Static motif images were incorporated to serve as reference points for anomaly detection, providing independent time range visualization. Finally, I developed a Python script that dynamically calculates anomalies ("discords") for each data sequence within Grafana, ensuring their easy identification.

3.2. Skills Developed

- **Data Analysis and Visualization:** The internship significantly bolstered my skills in data manipulation, employing advanced techniques for motif and discord identification. Additionally, I honed my ability to craft effective visualizations using industry-standard tools like Grafana.
- **Technical Skills:** Exposure to IBA Analyzer and the development of Python scripts for data analysis significantly expanded my technical skillset.
- **Problem-Solving:** Throughout the internship, I actively addressed challenges like identifying optimal thresholds, resolving configuration issues during dashboard import with Grafana, and automating discord highlighting, demonstrating strong problem-solving abilities.

3.3. Challenges Faced

- **Data Decoding and Processing:** Transforming raw mill data into a usable format presented a challenge, requiring a deep understanding of data processing techniques and careful attention to detail.
- **Threshold Optimization:** Finding the optimal thresholds for effectively identifying motifs and discords proved to be a challenge, demanding a balance between capturing meaningful patterns and avoiding oversensitivity.
- **Dashboard Enhancements:** Implementing static motif images and automating discord highlighting within Grafana presented unique challenges, requiring a blend of creativity and technical expertise to ensure seamless integration and user experience.

3.4. Achievements and Contributions

- **Improved Data Access and Analysis:** My data decoding efforts directly improved the accessibility and analysis potential of critical mill data, facilitating further process optimization.

- **Motif and Discord Identification:** I played a vital role in establishing a system to identify recurring patterns ("motifs") within the data. The inclusion of motifs in case studies allows for a more comprehensive understanding of process variations, potentially leading to improved control strategies.
- **Enhanced Data Visualization:** The creation of more informative and visually appealing dashboards through binary signal visualization, static motif images, and dynamic discord highlighting stands as a testament to my contributions, enhancing data exploration and communication.

3.5. Learning Outcomes

- **Industry Insights:** The internship provided valuable insights into the milling process and the dynamics of the industry, broadening my knowledge base and enhancing my understanding of practical applications.
- **Organizational Dynamics:** Direct observation provided a window into the organization's work culture, communication practices, and project management styles, offering valuable perspectives on professional workplace environments.
- **Personal and Professional Growth:** This experience significantly contributed to my personal development in areas like time management and communication. Furthermore, it has solidified my career aspirations within this field by providing practical exposure to the industry and its challenges. I further solidified my learning by utilizing the developed dashboards in a real-world setting at the BSP mill, gaining valuable insights into their effectiveness for data exploration and communication.

CHAPTER – IV

METHODOLOGY

CHAPTER IV

METHODOLOGY

This section details the methodology employed throughout the internship project.

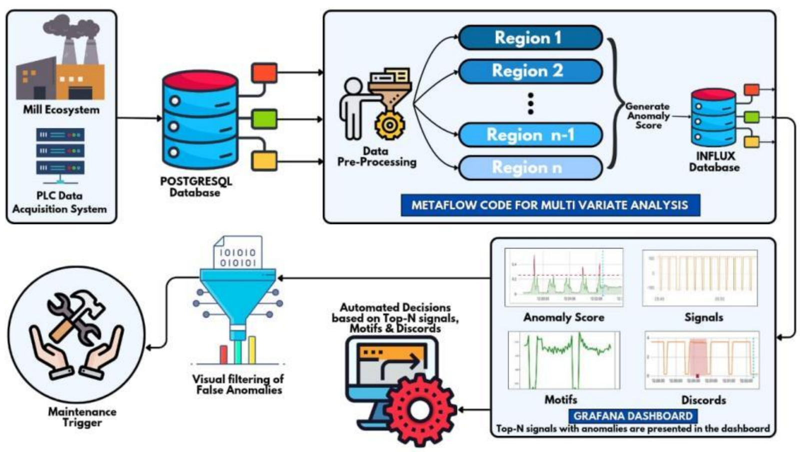


Fig 4.1: System Architecture

4.1 Project Overview

The internship project developed a system for identifying recurring patterns ("motifs") and deviations ("discords") within mill data. Data from the mill was decoded for training and directly used from the server in deployment. It was preprocessed, stored in InfluxDB, visualized on a Grafana dashboard, and used to generate alerts for timely anomaly detection. The objective was to enhance process optimization and anomaly detection capabilities. The project scope encompassed the following key phases:

- **Data Decoding:** Transforming raw mill data from a complex, proprietary format (.dat) into a readily usable format for analysis using IBA Analyzer software.

MONTH	DAYS
January	29 - 31
February	03 - 29
March	01 - 10, 23 - 31
April	01 - 20

Table 4.1.1: Data Decoded for the Specified Months

- **Motif and Discord Analysis:** Developing a system to identify recurring patterns ("motifs") and deviations ("discords") within the time series data.

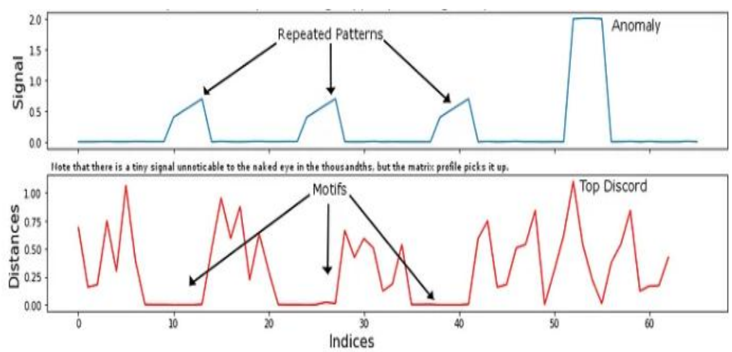


Fig 4.1.1: Illustration of Motifs

- **Dashboard Enhancements:** Creating informative and visually appealing Grafana dashboards to effectively communicate the identified motifs, discords, and other relevant data insights.

- **Real-World Deployment:** Deploying the developed dashboards at the BSP mill to facilitate real-time process monitoring and anomaly detection by plant personnel.

4.2 Research Design or Approach

The research design adopted a mixed-methods approach, combining quantitative techniques for data analysis with qualitative insights from domain experts. Time series data from the mill was analyzed using the following techniques:

4.2.1 Motif Identification:

- **Z-normalized Euclidean Distance:** This technique was employed to measure the similarity between different data segments, aiding in motif identification. The z-normalization step ensured all data points had a mean of 0 and a standard deviation of 1, allowing for fair comparison across different sensor values.
- **Matrix Profile Calculations:** This method identified recurring subsequences within the data, further solidifying the presence of motifs.

4.2.2 Discord Detection:

- **Euclidean Distance:** This technique measured the distance between data points to identify significant deviations from the established motifs.
- **Dynamic Time Warping (DTW):** This algorithm addressed the challenge of time series variations by warping one time series to match the other, allowing for more robust discord detection despite potential time shifts or scaling differences.

Threshold optimization was a crucial step, ensuring effective differentiation between anomalies ("discords") and normal process variations. This involved careful calibration to minimize false positives (incorrect anomaly flags) and false negatives (missed anomalies).

4.3 Data Collection Methods

The primary data source for this project was raw mill data collected from various sensors throughout the milling process. This data, obtained in a complex, proprietary format (.dat), required transformation for further analysis. To unlock the data's potential, we employed specialized software tools to convert it into a more accessible format (e.g., CSV) suitable for integration with our chosen data analysis tools. This transformation process ensured the data was readily interpretable, paving the way for in-depth analysis and the identification of crucial patterns for anomaly detection.

4.3.1 Data Decoding Techniques

Data decoding involved transforming the raw mill data from its original format (.dat) into a usable format for analysis within IBA Analyzer software. This process entailed:

- **Data Extraction:** Using appropriate tools or scripts, the relevant data segments encompassing the desired timeframes (e.g., January - April) were extracted from the .dat files.
- **Data Transformation:** The extracted data underwent necessary transformations to convert it into a compatible format with IBA Analyzer.
- **Data Validation:** The transformed data was meticulously checked for accuracy and consistency to ensure its suitability for further analysis.

4.4 Data Analysis Techniques

Following successful data decoding, the analysis phase employed the aforementioned techniques for motif and discord identification:

4.4.1 Motif Identification

Z-normalized Euclidean distance and matrix profile calculations were used to identify recurring patterns ("motifs") within the time series data.

4.4.1.1 Z-Normalized Euclidean Distance

- Z-Normalized Euclidean Distance is a variant of the Euclidean distance metric utilized in time series analysis.
- The formula for this is:

$$\mathbf{d_z(X, Y)} = \sqrt{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\frac{\mathbf{X_i - \mu_X}}{\boldsymbol{\sigma}_\mathbf{X}} - \frac{\mathbf{Y_i - \mu_Y}}{\boldsymbol{\sigma}_\mathbf{Y}} \right)^\mathbf{2}}$$

- (μ_X) and (σ_X) are the mean and standard deviation of vector ($\boldsymbol{X}$), respectively.
- (μ_Y) and (σ_Y) are the mean and standard deviation of vector ($\boldsymbol{Y}$), respectively.
- (n) is the number of elements in the vectors.

- It incorporates normalization by subtracting the mean and dividing by the standard deviation for each time series before computing the Euclidean distance.
- This normalization step helps to mitigate the impact of variations in scale and magnitude across dimensions, making it particularly useful for comparing time series with differing characteristics.

4.4.1.2 Matrix Profile

The matrix profile, generated by the STUMPY library, plays a crucial role in anomaly detection. This profile is a matrix that stores a distance value for each subsequence compared within the time series data. Lower values in the matrix profile correspond to motifs, representing recurring patterns observed throughout the data. Conversely, higher values indicate discords, which are deviations from these typical patterns and potentially signify anomalies in equipment operation.

```
Matrix profile for sensor motifs [1:17]:
[1.5515433048747542 1.5515433048747542 1.556903355981425]
Matrix profile for sensor motifs [1:19]:
[1.24279219039535 1.24279219039535 1.2540462980543843]
Matrix profile for sensor motifs [1:21]:
[2.354547764123694 2.354547764123694 2.3792590736559456]
Matrix profile for sensor motifs [1:36]:
[0.003682188159161121 0.003682188159161121 0.0036927104770469346]
Matrix profile for sensor motifs [1:40]:
[0.5796665150146385 0.5796665150146385 0.582189692664842]
Matrix profile for sensor motifs [1:71]:
[5.1588643828449126 5.1588643828449126 5.2056828501081425]
```

Fig 4.4.2.1.1 Matrix Profile for Sensors Motifs

The matrix profile, generated for various sensors, offers a valuable tool for identifying potential anomalies. This matrix stores distance values, where lower values indicate recurring patterns (motifs) observed in the sensor data. Conversely, higher values signify deviations (discords) from these patterns and warrant further investigation.

Within the matrix, the lowest values (0 in this case) represent the most prominent motifs, while the highest value (infinity) indicates the greatest deviation from normal behavior.

```
Matrix profile for sensor motifs [1:21]:

[2.354547764123694 2.354547764123694 2.3792590736559456]
```

Fig 4.4.2.1.2: Matrix Profile value for Signal [1:21]

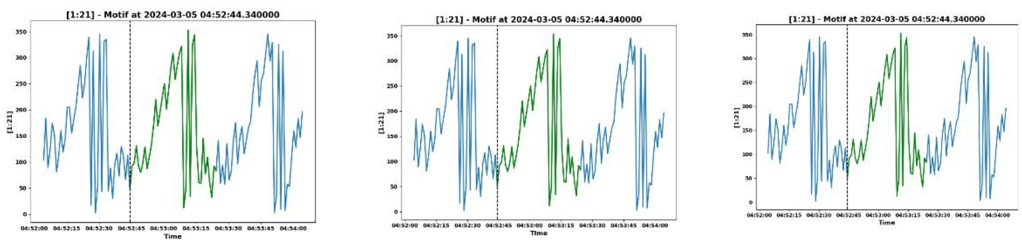


Fig 4.4.2.1.3: Top 3 Motifs for Signal [1:21]

The provided images showcase two key aspects of anomaly detection: the matrix profile itself and the corresponding sensor data sequences. The low values observed in the matrix profile visually confirm the presence of recurring patterns, evident in the plot of the sensor readings. These consistent patterns within the data qualify them as motifs, representing the typical behavior of the sensor under normal operating conditions.

4.4.2: Discord Detection

Euclidean distance combined with Dynamic Time Warping (DTW) was employed to detect deviations ("discords") from the established motifs. These discords potentially represent anomalies within the milling process.

DTW Distance

- Dynamic Time Warping (DTW) distance is a specialized measure used to quantify the similarity between two time series sequences, accommodating variations in their lengths or speeds by allowing for optimal alignment through warping or stretching of the time axis.

- The formula for this is:

$$\mathbf{wps[i,j] = (s1[i] - s2[j])2 + min(wps[i-1,j] + penalty, wps[i,j-1] + penalty, wps[i-1,j-1])}$$

  - i,j: Indices representing positions within the two time series sequences being compared.
  - $s1[i], s2[j] s1[i], s2[j] s1[i], s2[j]$: Values from the respective time series sequences at positions iii and jjj.
  - penalty: Penalty value for introducing gaps (insertions or deletions) during the warping process.
  - $wps[i-1,j]$, $wps[i,j-1]$, $wps[i-1,j-1] wps[i-1,j]$, $wps[i,j-1]$, $wps[i-1,j-1] wps[i-1,j], wps[i,j-1], wps[i-1,j-1]$: Accumulated costs of aligning up to positions $(i-1,j), (i,j-1), (i-1,j), (i,j-1), (i-1,j), (i,j-1)$, and (i−1,j−1)(i-1, j-1)(i−1,j−1) in the warping path matrix.
  - $dtw = wps[-1,-1] dtw = \sqrt{wps[-1,-1]} dtw = wps[-1,-1]$: Final Dynamic Time Warping distance between the two sequences (square root of the accumulated cost at the end of the warping path.

- Unlike conventional distance metrics, DTW considers local variations in sequences, making it particularly effective for comparing time series data with temporal distortions, shifts, or irregularities.
- Widely applied in domains like speech and pattern recognition, DTW captures both local and global similarities between sequences, offering a robust measure of similarity that is insensitive to temporal variations.
- It has a range of 0 -1. Any score lesser than 0.5 indicates similarity.

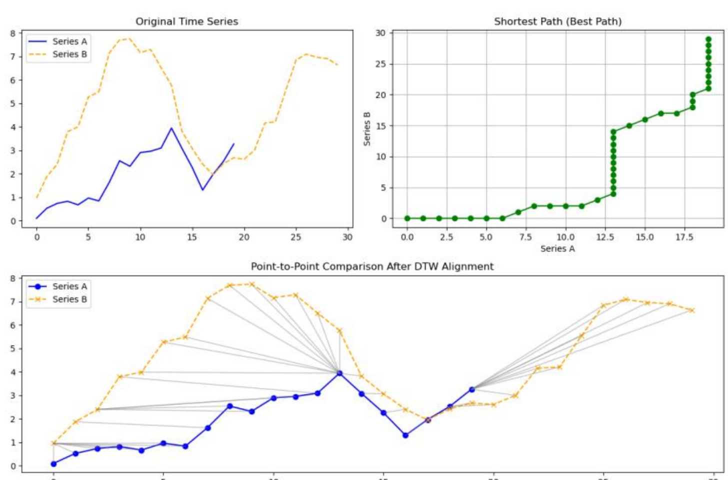


Fig 4.4.2.1: Working of DTW Distance

Results/ Key Findings:

To evaluate anomaly detection techniques, Euclidean and Dynamic Time Warping (DTW) distances were compared across various sensors. Here, a sensor's motif plot, sequence, and discord score are shown. This comparison will assess which distance measure better identifies deviations from the sensor's typical behavior (motif).

Sensor: [9:17]_Stand 18_Torque

Fig 4.4.2.2: Original and Motif Plot

Metric	***Score***	***Range***
Euclidean Distance	13.27	0 - ∞
DTW Distance	0.315	0 – 1

Table 4.4.2.1: Results of Euclidean and DTW Distance

Euclidean distance vs DTW distance for discords

- Euclidean distance may mistakenly flag more discords due to its sensitivity to outliers and variations, increasing false positives.
- It fails when sequences have zero standard deviation, limiting its applicability in highly uniform data.
- Z-Normalized Euclidean requires equal-length sequences, unlike DTW, which handles sequences of varying lengths flexibly.

Following analysis, Dynamic Time Warping (DTW) distance was chosen for its ability to capture similarity in time-varying sequences. This allows for effective anomaly (discord) detection by comparing DTW distance to a predefined threshold (currently 0.8). Sequences exceeding the threshold are classified as discords, indicating significant deviations from the expected pattern represented by the motif. Additionally, a comprehensive evaluation will be conducted using a diverse set of benchmark datasets to validate the robustness and generalizability of the DTW-based anomaly detection framework. By leveraging DTW distance, the proposed method provides a powerful mechanism for identifying unexpected patterns.

4.5 Ethical Considerations

Throughout the project, ethical considerations were paramount. Data privacy was strictly maintained, ensuring all data handling adhered to relevant regulations and institutional policies. Anonymization techniques were employed to further safeguard sensitive information, and user consent was obtained whenever necessary. This commitment to ethical data practices fosters trust and transparency, crucial for the long-term success of such projects.

4.6 Limitations

The limitations include the frequent changes in motifs for signals, which complicate the accurate interpretation of these signals. This variability requires constant adjustments and fine-tuning, adding a layer of complexity to the process. Additionally, the large scale of the mill presents significant challenges in determining which fluctuations are critical and which are inconsequential. The vastness of the mill means that even minor fluctuations can have varying impacts, making it difficult to establish clear thresholds for action. Consequently, this requires a sophisticated monitoring system and nuanced analysis to ensure accurate decision-making.

CHAPTER – V

IMPLEMENTATION

CHAPTER V

IMPLEMENTATION

This section delves into the project's implementation, detailing the system developed to identify recurring patterns ("motifs") and deviations ("discords") within mill data. This system plays a vital role in enhancing process optimization and anomaly detection capabilities. By analyzing the data for these patterns and deviations, the system aims to improve equipment performance, ultimately contributing to increased efficiency and reduced downtime.

5.1 Development Environment

This project utilized the following software and tools:

Software Tools:

- **Python:** A versatile programming language widely used for scientific computing, data analysis, and machine learning. In this project, Python served as the foundation for developing the motif and discord identification system.
- **NumPy:** A fundamental Python library for numerical computations. NumPy played a crucial role in various aspects of time series data analysis, including:
- **stumpy:** A Python library designed for scalable and efficient time series motif and discord discovery. Stumpy provides a comprehensive suite of algorithms for analyzing time series data, particularly focusing on identifying repeated patterns (motifs) and anomalies (discords) within large datasets.
- **dtaidistance:** This library provides various distance metrics specifically designed for time series data. It supported calculations like Euclidean distance used in motif identification.
- **softdtw (Dynamic Time Warping):** This library implements the Dynamic Time Warping (DTW) algorithm, crucial for robust anomaly (discord) detection in time series with potential time shifts or scaling variations.

- **InfluxDB:** A time series database optimized for storing and querying large volumes of time-stamped data. In this project, InfluxDB served as the repository for storing the mill data.
- **Grafana:** A popular open-source visualization tool for creating interactive dashboards. Grafana dashboards were used to communicate the identified motifs, discords, and other insights gleaned from the mill data analysis.

Hardware: General-purpose computer

Configuration

This section outlines the steps to set up the system for deployment:

1. InfluxDB Database Setup:

- Install InfluxDB according to the official documentation.
- Create a database specifically for storing the mill data.
- Define data retention policies (optional) to manage storage based on your needs.
- Secure your InfluxDB instance using appropriate authentication and authorization mechanisms.

2. Grafana Setup:

- Install Grafana according to the official documentation.
- Configure a data source pointing to your InfluxDB instance, specifying the database name and credentials.
- This data bucket serves as the primary repository for all data utilized by the dashboard visualizations.
- Import the pre-configured Grafana dashboard JSON file. This file should contain the visualizations and configurations for displaying motifs, discords, and other relevant data.

3. Configuration Adjustments:

Update the dashboard configuration to reflect in specific environment, such as:

- Modifying the InfluxDB data source name and credentials to match setup.
- Replacing placeholder values in the dashboard queries with the actual bucket names or data field names used in InfluxDB database.
- Adjusting user permissions within Grafana to control access to the dashboards.

5.2 Project Execution

The project was executed in the following phases:

Data Acquisition and Preprocessing

- Raw mill data in .dat format was obtained from the BSP mill.
- **Data Decoding:** Specialized tools or scripts were used to extract relevant data segments for the desired timeframes (e.g., January - April) from the .dat files.
- **Data Transformation:** The extracted data underwent transformations like unit conversions, scaling adjustments, and time-series formatting to make it compatible with IBA Analyzer software.
- **Data Cleaning and Validation:** The data was meticulously checked for accuracy and consistency to ensure its suitability for further analysis.
- **Potential Preprocessing with stumpy:** Stumpy's functionalities for normalization, smoothing, or segmentation might have been applied to prepare the data for motif discovery.

Motif Discovery and Window Size Analysis

- **Initial Motif Identification:** Motifs (recurring patterns) were identified within the time series data for various sensors using techniques like stumpy's Matrix Profile algorithm.
- **Window Size Dependence:** An important discovery was the dependence of identified motifs on the chosen window size (e.g., 40, 60, 80, 100 seconds). This analysis was performed for all sensors and window sizes.

- **Profile Dependence:** A further layer of complexity was revealed as motifs were found to be dependent on the specific mill profile (BRM Mill profiles: 10mm, 12mm, 16mm, 20mm). Extensive analysis across various sensors was conducted to determine the optimal motif for each window size and profile combination.
- **True/False Positive Analysis:** To identify the window size that captured the most variation aiding in anomaly detection, an analysis of true and false positives was conducted. This evaluation led to the conclusion that an 80-second window size yielded the most desirable results across all profiles.

Discord Detection and Distance Metric Selection

- **Distance Metric Exploration:** Candidate distance metrics for discord (anomaly) detection were explored, including z-normalized Euclidean distance and Dynamic Time Warping (DTW) distance.
- **DTW Library Selection:** Different libraries implementing DTW were evaluated for performance, with the focus on identifying the fastest option for real-time application.

Grafana Dashboard Development

Challenges were encountered in displaying motifs within the Grafana dashboard due to the dynamic range of the data. This involved exploring various techniques to effectively represent the motifs as the range changed.

Deployment and Monitoring:

The final Grafana dashboards were integrated into the BSP mill's monitoring system, facilitating real time process visualization.

5.3 Timeline

The project execution spanned approximately four months, with the following key milestones:

Month 1:

Focused on data acquisition and initial preprocessing. This involved:

- Obtaining raw mill data in .dat format from the BSP mill.
- Performing data cleaning and validation to ensure data quality.
- Applying data, utilizing stumpy for motif discovery.
- Identifying and extracting relevant features from the sensor data.

Month 2:

Delved into motif discovery and window size analysis. During this phase:

Algorithms like stumpy's Matrix Profile were implemented to identify initial motifs within the sensor data. A systematic analysis was conducted to explore the dependence of these motifs on the chosen window size (e.g., 40, 60, 80, 100 seconds).

Weeks 3-4 focused on:

Investigating the influence of specific mill profiles (10mm, 12mm, 16mm, 20mm) on the identified motifs.

Conducting an analysis of true and false positives to determine the window size that captures the most variation for anomaly detection.

Month 3:

Tackled discord detection and distance metric selection. This month involved:

- Exploring candidate distance metrics like z-normalized Euclidean distance and Dynamic Time Warping (DTW) for discord (anomaly) detection.
- Evaluating different DTW libraries for performance, prioritizing speed for real-time application.

Weeks 3-4 concentrated on:

- Designing and enhancing dashboards in Grafana to effectively visualize identified motifs, discords, and other relevant data.
- Addressing challenges in presenting motifs due to the dynamic range of the data.

Month 4:

Focused on monitoring, with a primary emphasis on evaluating the effectiveness of the implemented motif and discord approach. This involved:

- Integrating the final Grafana dashboards into the BSP mill's monitoring system.
- Initiating monitoring of the dashboards by designated personnel.

- Developing case studies to analyze the system's performance in real-world scenarios, focusing on how well the identified motifs and discords aided in anomaly detection.
- Analyzing the effectiveness of the chosen window size and distance metrics based on the case studies.

5.4 Resource Allocation

The project primarily involved my time and expertise dedicated to data analysis, algorithm development, and dashboard creation. Additionally, collaboration with a mentor likely provided valuable guidance and domain knowledge.

One person was designated to monitor the dashboards and develop case studies to analyze the effectiveness of the implemented system.

5.5 Challenges Faced

- **Data Preprocessing and Feature Engineering:** While not explicitly mentioned, data preprocessing and feature engineering are common challenges in time series analysis. Extracting meaningful features from raw sensor data can be an iterative process.
- **Window Size and Profile Dependence of Motifs:** Discovering the dependence of motifs on window size and mill profiles added complexity to the analysis and required additional investigation.
- **Balancing True/False Positives for Anomaly Detection:** Finding the optimal balance between true and false positives in anomaly detection is crucial for avoiding missed anomalies or overwhelming alerts.
- **Grafana Visualization Challenges:** Effectively representing motifs within the dashboard given dynamic data ranges presented a hurdle that required exploring alternative visualization techniques.

5.6 Success Factors

- **Mentor Expertise and Communication:** Effective communication and guidance from mentor played a significant role in the project's success. Their domain knowledge and experience would have been invaluable throughout the process.
- **Teamwork:** Collaborative efforts, potentially with your mentor or other colleagues, could have contributed to problem-solving and achieving project goals.
- **Community Forums and Documentation:** Utilizing online communities, forums, and technical documentation likely proved beneficial for troubleshooting challenges, learning about best practices, and finding relevant libraries and algorithms.

5.7 Lessons Learned

The project provided valuable learning experiences in several areas:

- **Time Series Analysis:** The project deepened understanding of time series data analysis techniques, including motif discovery, anomaly detection using distance metrics, and the importance of data preprocessing for accurate results.
- **Distance Metrics:** The project explored the application of distance metrics like Euclidean distance and Dynamic Time Warping (DTW) for assessing similarities and deviations in time series data. DTW's ability to handle time shifts proved advantageous in anomaly detection.
- **Human-in-the-Loop Approach for Improved Accuracy:** A key takeaway was the effectiveness of a human-in-the-loop approach to enhance accuracy. While existing machine learning models like LSTMs might have yielded lower precision, the motif-discord approach combined with human monitoring offered a way to iteratively improve accuracy. This highlights the value of human expertise in conjunction with machine learning models for certain tasks.

This approach allows human analysts to leverage their domain knowledge and judgment to refine the system's performance over time. The project demonstrates that even with advanced models like LSTMs, human oversight can be crucial in real-world applications.

CHAPTER – VI

RESULTS AND DISCUSSION

CHAPTER VI

RESULTS AND DISCUSSION

6.1 Presentation of Results

The following sections present the key outcomes and their broader significance.

6.1.1 Binary Signals Visualization

To enhance interpretability of digital (binary) signals prevalent across various system regions, a novel visualization approach was developed. This method leverages signal naming conventions to identify binary signals and subsequently transforms them into bar charts. The bar chart representation directly correlates bar length with the duration of each binary state (0 or 1), enabling swift and accurate visual comparison of signal widths. This innovation streamlines data analysis and decision-making processes.

Fig 6.1.1: Binary Signals Width Plotting

The figure presents a two-part visualization of the binary signals. The top panels display the original plots of the signals. These plots allow us to directly observe the transitions between the two states (0 and 1) over time.

The bottom panels depict the distribution of signal widths for both states (0 and 1). This visualization is typically presented as a bar chart, where the x-axis represents the duration (width) of each state and the y-axis represents the frequency of occurrence for each duration.

To enhance visual interpretation, the bars within the chart are typically color-coded. Red bars commonly represent the width of the "1" state, while green bars represent the width of the "0" state. This color scheme facilitates the quick identification and comparison of the duration for each signal state.

6.1.2 Motif Integration for Anomaly Detection

A crucial aspect of the dashboard enhancement involved the incorporation of motifs. Motifs, representing generalized, repetitive patterns within the signal behavior, are displayed alongside the original signal plot. This facilitates a comparative analysis, allowing users to readily identify deviations from the expected pattern. This visual anomaly detection approach empowers users to make informed judgments regarding the normalcy of signal behavior, enhancing overall system monitoring effectiveness.

Initial Approach: Dynamic Motifs in InfluxDB

My initial approach involved storing the data in InfluxDB and directly displaying the motifs within the dashboard by fetching data on demand. However, this method presented a challenge: the motifs would disappear whenever the user adjusted the dashboard's time range.

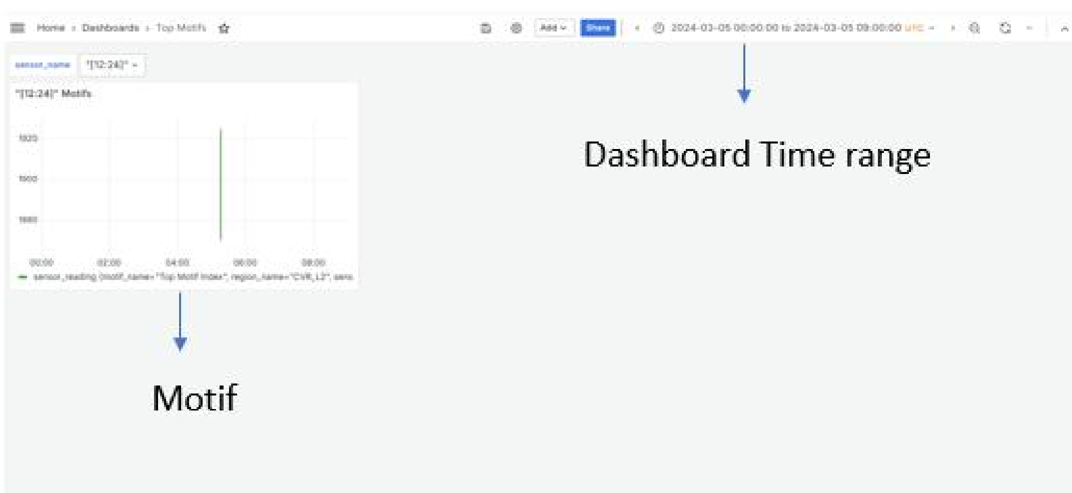


Fig 6.1.2.1: Motifs dependent on dashboard time range

As illustrated in Figure 6.1.2.1, the motifs were dynamically generated based on the current time window displayed in the dashboard. This resulted in the motifs becoming unavailable when the user changed the time range.

But we need motifs to be static and independent of the dashboard range, for this I tried several approaches but none worked. Grafana currently does not support time locking or static rendering or dynamic time fetching.

Challenges and Alternative Solutions

Static Motifs are Required: For effective anomaly detection, the motifs need to be static and independent of the dashboard's time range. This allows for a consistent reference point for comparison with the actual signal data.

Exploring Alternative Approaches: To achieve static motif visualization, I investigated several methods, but none proved successful. Currently, Grafana lacks built-in functionalities for "time locking," "static rendering," or "dynamic time fetching" that would enable this desired behavior. Grafana's current functionalities, such as relative time options ("1d" or "1h"), still rely on a dynamic time window. Unfortunately, these features don't support locking the time range to a specific static interval, which was my desired behavior for motif visualization.

Exploring Alternative Monitoring Systems: Exploring Prometheus, another monitoring system, revealed integration difficulties due to data model incompatibilities. Additionally, limitations in pushing data specific to this project's needs, along with the lack of static time range locking similar to Grafana, hindered a successful implementation for static motif visualization.

Solution: Motif Image Addition

To address the challenge of dynamic motifs disappearing with time range changes, a solution involving pre-generated static images was implemented.

Motif Generation:

Motifs were first calculated for each profile. This involved analyzing the signal behavior within each profile to capture the characteristic repetitive patterns.

Image Storage with Organized Structure:

To ensure compatibility with file naming conventions, folders were created for each profile. Within each folder, the motifs were saved as images with descriptive names. This naming approach avoided using IDs that might contain special characters like ":". These characters can cause issues with file systems.

Motif Integration in Dashboard:
The final step involved integrating these pre-generated motif images into the dashboard. By displaying the appropriate motif image alongside the corresponding sensor data, users can visually compare the actual signal behavior with the expected pattern represented by the motif.

Image Fetching and Display: Based on the extracted motif name, the associated image file (stored with descriptive names in profile folders) was retrieved. Finally, the retrieved motif image was displayed within the dashboard alongside the corresponding sensor data.

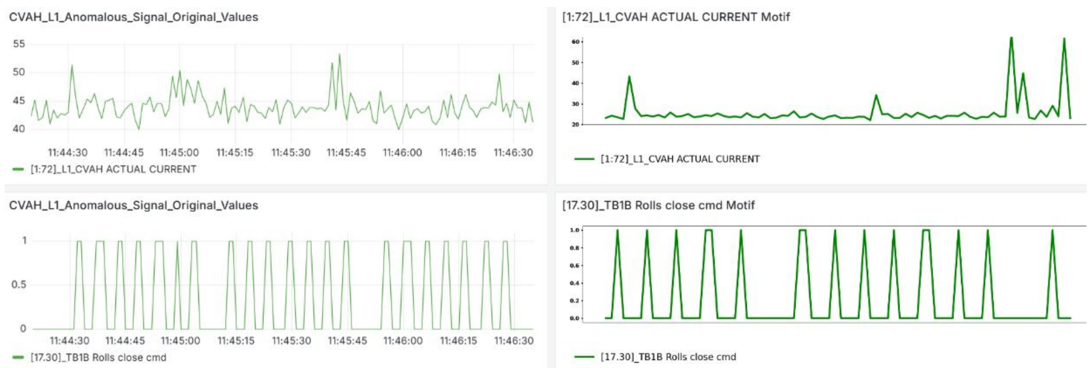


Fig 6.1.2.2: Motifs Images independent of dashboard time range

6.1.3 Dynamic Discord Highlighting with Python Script

To achieve real-time anomaly detection within the dashboard, a custom Python script was implemented. This script dynamically calculates and tags anomalies (discords) within data sequences. This approach offers several advantages:

1. **Real-time Detection:** Discords are identified as the data streams, ensuring continuous anomaly monitoring.
2. **Flexibility and Customization:** The Python script provides a versatile environment for implementing various anomaly detection algorithms, allowing for tailored solutions based on specific project requirements.
3. **Enhanced Visualization:** Anomalies are dynamically highlighted within the dashboard, enabling clear visual distinction between normal data and potential issues, facilitating rapid response and troubleshooting.

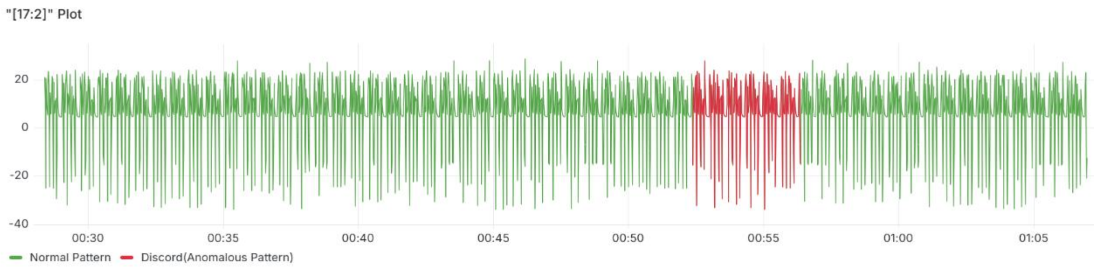


Fig 6.1.3.1: Discords highlighting in Grafana

This figure represents the sensor pattern, the red highlighted part is discord and the other part is normal pattern of sensor.

6.1.4 Case Studies

6.1.4.1 True Positive Case Studies

This analysis of 11 cobble/anomaly events details the date, time, and specific reason (identified as "Failure type") for each failure, along with the affected region, mill profile, potentially causative signals, and the time these signals deviated from normal. By examining these factors across all cases, we can identify trends, delve deeper into root causes, and potentially develop a system to predict and prevent future cobble events by analyzing signals and lead times.

Case Study 1: COBBLE IN STAND 18

- During the COBBLE IN STAND 18 + PLAN JOB, with a 12MM profile, monitoring Stand_7_12 signals, sensor [9:37]_Stand 08_RPM maintained a consistent pattern until approximately 08:50, after which it experienced a sudden drop.
- The same pattern was observed for other signals, such as [9:34]_Stand 07_LineSpeed and [9:42]_Stand 10_RPM, where they also abruptly decreased at around 08:50 before resuming an increasing trend.
- During analysis of COBBLE IN STAND + PLAN JOB (12mm profile) data, a synchronized decrease in both line speed and RPM (sensor readings) was observed around 08:50. This deviation from normal operational patterns suggests a potential anomaly within the production system.

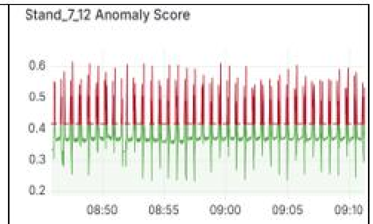
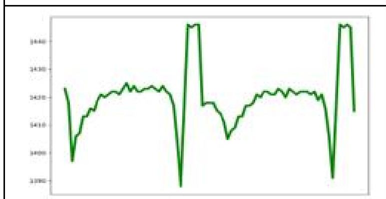
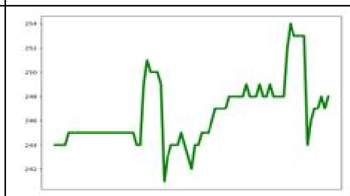
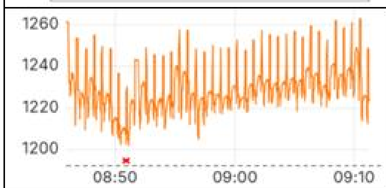
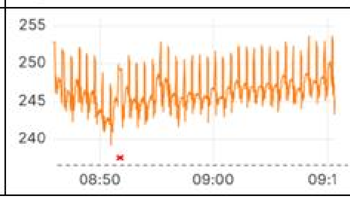
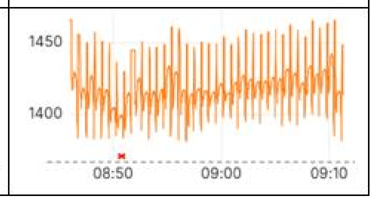
Actual cobble: 2024-06-03 08:52 - 2024-06-03 10:53 IBA time: 2024-06-03 09:05 - 2024-06-03 11:06 (Iba Time 13 minutes) COBBLE IN STAND 18 + PLAN JOB PROFILE: 2*12MM	Considering Stand_7_12 Signals Lead Time: 13 min	
[9:37]_Stand 08_RPM	**[9:34]_Stand 07_LineSpeed**	**[9:42]_Stand 10_RPM**
		
		

Table 6.1.4.1.1: Cobble in STAND-18

The table shows the main sensor plots and the motif plots, revealing that these patterns deviate significantly from each other. This high degree of deviation indicates that the current sensor behaviors are not aligning with the expected normal patterns, signifying the presence of discords. Discords are instances where the actual sensor readings differ substantially from the expected patterns, often pointing to anomalies or irregularities in the system.

Case Study 2: COBBLE LINE-1 FFB

- During the cobble line-1 FFB + plan job, with a 12MM profile, monitoring PINCHROLL_L1 signals, sensor [1:22]_L1_PR2_ActSpeed maintained a regular pattern until approximately 06:55, after which it began to fluctuate.
- At 06:58, sensor [1:69]_L1_PR2_ACTUAL_CURRENT experienced a sudden decrease.
- Concurrently, sensor [1:24]_L1_PR2_ActTorque showed a decline in value, commencing at the same time as sensor [1:69]_L1_PR2_ACTUAL_CURRENT.
- In summary, the regular operation indicated by the speed sensor until 06:55 was disrupted, leading to fluctuating speed readings.

- This was followed by a sudden decrease in both current and torque at 06:58, suggesting a significant change in the operating conditions of the pinch roll system, likely due to mechanical or electrical issues.

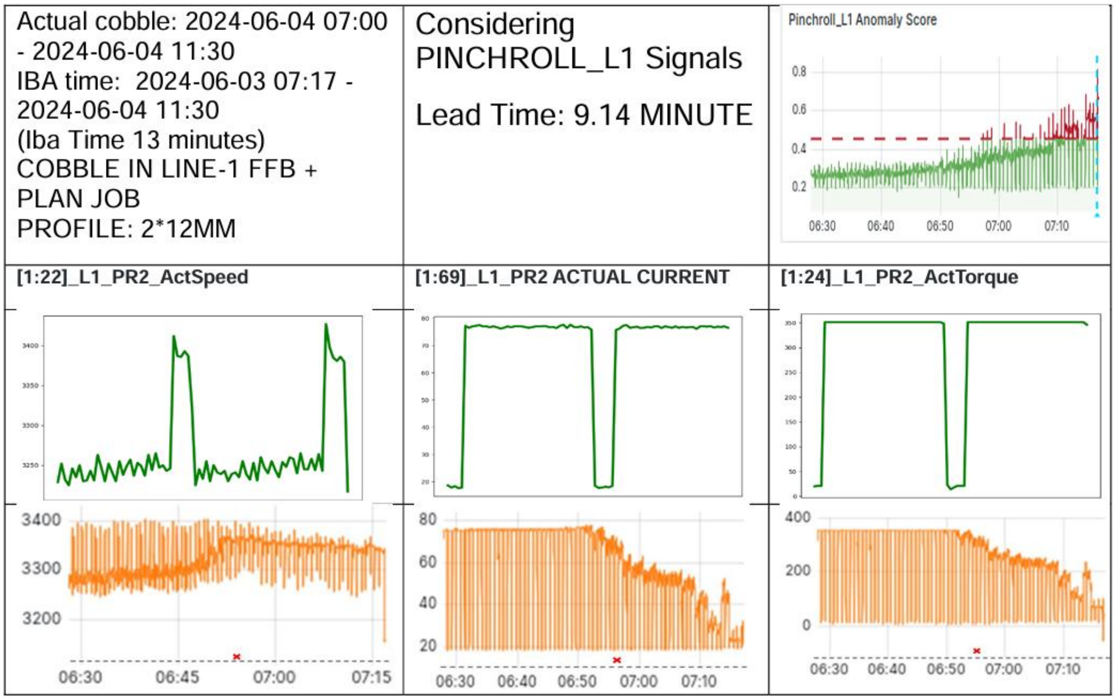


Table 6.1.4.1.2: Cobble in LINE-1 FFB

The table shows the main sensor plots and the motif plots, revealing that these patterns deviate significantly from each other. This high degree of deviation indicates that the current sensor behaviors are not aligning with the expected normal patterns, signifying the presence of discords. Discords are instances where the actual sensor readings differ substantially from the expected patterns, often pointing to anomalies or irregularities in the system.

Case Study Summary:

The **datetime** column indicates the precise time and date of the cobble incident. The **failure type** column specifies the nature of the failure that occurred. The **profile** column describes the mill's operational profile at the time of the cobble. The **region** column identifies the area affected by or involved in the cobble event. The **signals involved** column lists the primary signals implicated in the cobble. The **lead time** column denotes the duration within which we were able to predict the cobble before its occurrence.

Here is a comprehensive summary of all case studies, including the signals involved and the corresponding lead times.

Date time	Failure type	Profile	Region	Signals involved	Lead time
2024-03-06 12:51:43 - 2024-03-06 13:59:44	OVER SHOOT AT LINE-2	16 mm	CVR_L2, CVAH_L2, Pinch Roll-L2	[12:23]_FFB Line 2 Torque Act [12:24]_FFB Line 2 Line Speed [12:70]_FFB2 Motor Current [5:74]_L2_DCVAH_ Position error [5:24]_L2_PR4_ActTorque [5:65]_L2_PR4 Roll Closing Pressure	4.46 minutes
29-03-2024 03:43:10 — 29-03-2024 04:24:40	COBBLE AT WATER BOX LINE-2 FFB	20 mm	CVR_L2, Pinch Roll-L2	[5:29]_L2_PR1_ActTorque [5:80]_L2_DCVR_output current feedback from drive [9:77]_FFB Line 2 Loop Height [12:24]_FFB Line 2 Line Speed [5:27]_L2_PR1_ActSpeed [12:70]_FFB2 Motor Current	14.30 minutes
2024-03-29 08:04:34 - 2024-03-29 08:23:48	COBBLE IN LINE-1 FFB	20 mm	CVAH_L1, Pinchroll_L1	[17:13]_TWC 1A Act Pos [17:47]_TB1A Motor Current [1:70]_L1_PR3 ACTUAL CURRENT [1:60]_L1_PR2 Roll Closing Pressure [1:69]_L1_PR2 ACTUAL CURRENT [1:27]_L1_PR3_ActTorque	18.48 minutes

Table 6.1.4.1.3: True Positive Case Studies Summary

6.1.4.2 False Positive Case Studies

This analysis considers the possibility of false positives, where the system might have flagged an anomaly even though none occurred. Additionally, the case studies incorporate markings on anomalous patterns and utilize motifs to further aid in understanding these events.

Case Study 1

COBBLE INDICATOR TIME: 2024-04-05 01:51:16
PROFILE - 2*12MM

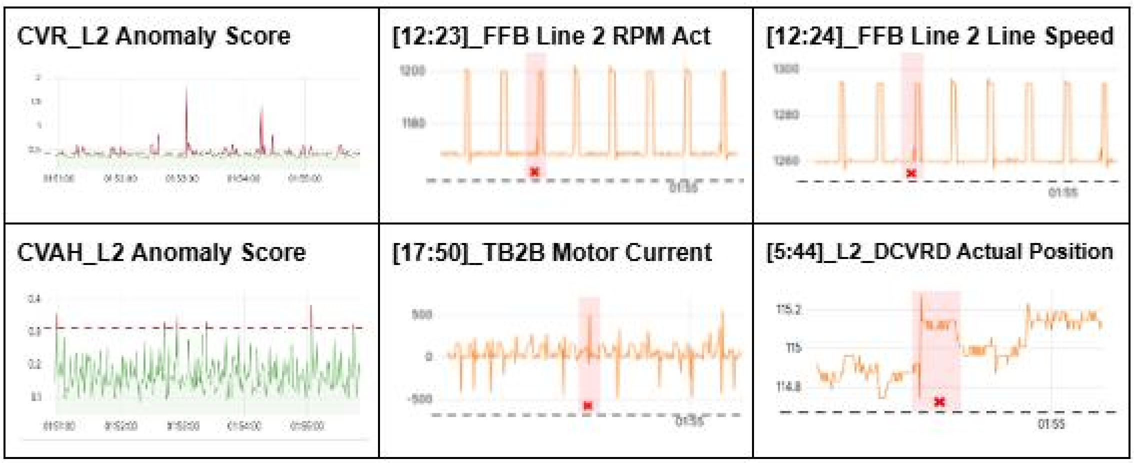


Table 6.1.4.2.1: False Positive Case Study 1

False positives occur when the system incorrectly predicts a cobble, typically due to changing patterns such as a sudden decrease in sensor values or erratic fluctuations. In this case study, a false positive was predicted in the CVAH_L1 and PINCHROLL_L2 region.

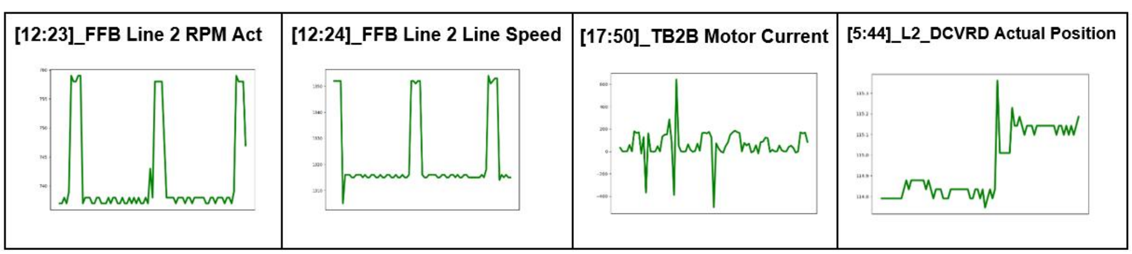


Table 6.1.4.2.2: Motifs for the Sensors for Case Study 1

CVAH_L1 Region Analysis: In the CVAH_L1 region, an anomaly score crossed the threshold at approximately 2024-04-15 13:25:09, with earlier occurrences at 2024-04-15 13:23:09.

Cobble was indicated at 2024-04-15 13:25:09. The signals involved are [1:59]_E35_CVAH_Actual Torque and [1:33]_L1_PR4_ActTorque.

- [1:59]_E35_CVAH_Actual Torque: This signal exhibited a constant pattern. However, at 2024-04-15 13:25:09, there was a slight deviation. Upon comparison with the standard motif (general pattern) and the normal pattern, this deviation appears typical. Therefore, no actual cobble occurred, and the model's indication is a false positive.

- [1:33]_L1_PR4_ActTorque: This signal also showed a constant pattern, with a slight increase in the pattern's lower width at 2024-04-15 13:25:09. Similar to the previous signal, when compared with the standard motif and the normal pattern, this deviation is deemed normal. Consequently, no actual cobble occurred, and the model's indication is a false positive.

PINCHROLL_L2 Region Analysis: In the PINCHROLL_L2 region, an anomaly score crossed the threshold at approximately 2024-04-15 13:19:58, with cobble indicated at 2024-04-15 13:23:09. The signals involved are [5:19]_L2_PR2_ActSpeed and [5:22]_L2_PR3_ActSpeed.

- [5:19]_L2_PR2_ActSpeed: The pattern for this signal remained constant between 3150-3350. However, at 2024-04-15 13:25:09, the value ranged between 3100-3255. Although the system predicted an anomaly, this deviation aligns with the normal pattern for this sensor and is thus considered typical.

- [5:22]_L2_PR3_ActSpeed: This signal exhibited a constant pattern between 3100-3300. Around 2024-04-15 13:25:09, there was a lower spike, which triggered the anomaly prediction. However, considering the standard motif, such spikes are normal for this signal within this range. Therefore, this deviation is acceptable.

In both regions, the deviations observed in the signal patterns were consistent with the standard motifs and normal patterns. Thus, the anomalies indicated by the model were false positives, with no actual cobble events occurring.

Case Study 2

COBBLE INDICATOR TIME: 2024-04-15 13:25:09
PROFILE - 2*12MM

False positives occur when the system incorrectly predicts a cobble, typically due to changing patterns such as a sudden decrease in sensor values or erratic fluctuations. In this case study, a false positive was predicted in the CVAH_L1 and PINCHROLL_L2 region. False positives in cobble prediction, particularly in sensitive regions like CVAH_L1 and PINCHROLL_L2, can cause significant disruptions.

Understanding the factors that lead to these erroneous predictions—such as sudden decreases in sensor values and erratic fluctuations—enables the development of strategies to mitigate their occurrence.

Through improved sensor accuracy, advanced data analysis, redundancy, and environmental control, the reliability of cobble predictions can be enhanced, leading to more efficient and effective operations.

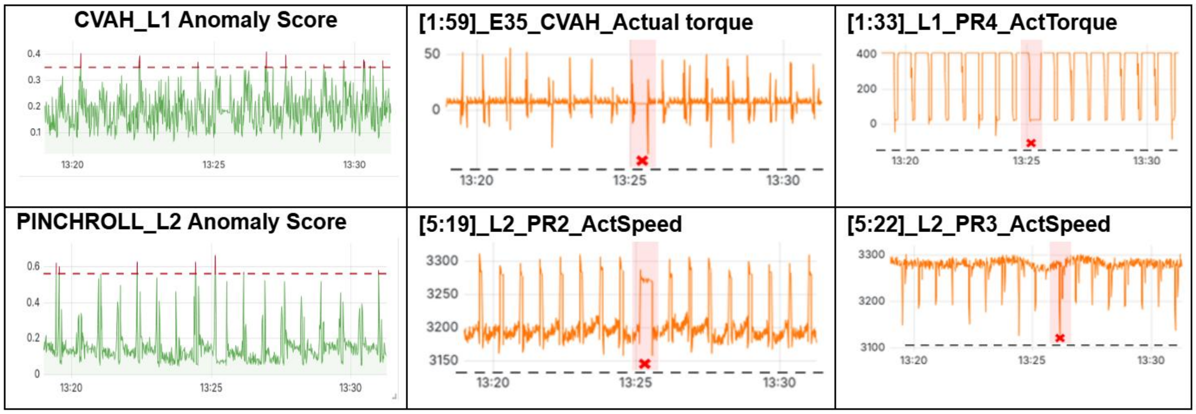


Table 6.1.4.2.3: False Positive Case Study 2

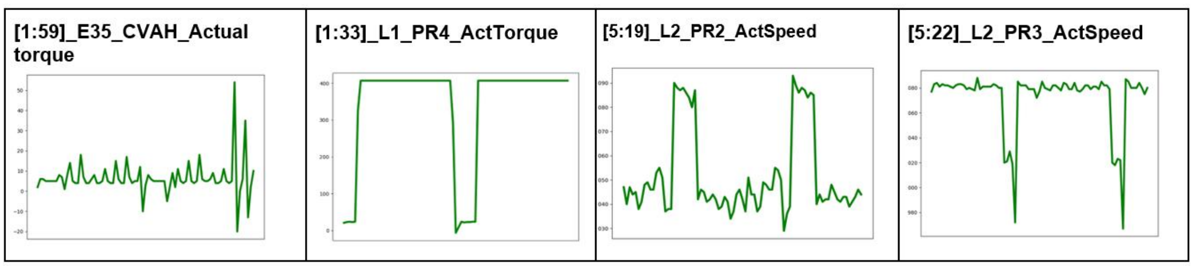


Table 6.1.4.2.4: Motifs for the Sensors for Case Study 2

CVAH_L1 Region Analysis: In the CVAH_L1 region, an anomaly score crossed the threshold at approximately 2024-04-15 13:25:09, with earlier occurrences at 2024-04-15 13:23:09. Cobble was indicated at 2024-04-15 13:25:09. The signals involved are [1:59]_E35_CVAH_Actual Torque and [1:33]_L1_PR4_ActTorque.

- [1:59]_E35_CVAH_Actual Torque: This signal exhibited a constant pattern. However, at 2024-04-15 13:25:09, there was a slight deviation. Upon comparison with the standard motif (general pattern) and the normal pattern, this deviation appears typical. Therefore, no actual cobble occurred, and the model's indication is a false positive.

- [1:33]_L1_PR4_ActTorque: This signal also showed a constant pattern, with a slight increase in the pattern's lower width at 2024-04-15 13:25:09. Similar to the previous signal, when compared with the standard motif and the normal pattern, this deviation is deemed normal. Consequently, no actual cobble occurred, and the model's indication is a false positive.

PINCHROLL_L2 Region Analysis: In the PINCHROLL_L2 region, an anomaly score crossed the threshold at approximately 2024-04-15 13:19:58, with cobble indicated at 2024-04-15 13:23:09. The signals involved are [5:19]_L2_PR2_ActSpeed and [5:22]_L2_PR3_ActSpeed.

- [5:19]_L2_PR2_ActSpeed: The pattern for this signal remained constant between 3150-3350. However, at 2024-04-15 13:25:09, the value ranged between 3100-3255. Although the system predicted an anomaly, this deviation aligns with the normal pattern for this sensor and is thus considered typical.

- [5:22]_L2_PR3_ActSpeed: This signal exhibited a constant pattern between 3100-3300. Around 2024-04-15 13:25:09, there was a lower spike, which triggered the anomaly prediction. However, considering the standard motif, such spikes are normal for this signal within this range. Therefore, this deviation is acceptable.

In both regions, the deviations observed in the signal patterns were consistent with the standard motifs and normal patterns. Thus, the anomalies indicated by the model were false positives, with no actual cobble events occurring.

Case Study Summary

Here's a summary of all the false positive case studies I did with the regions and signals involved.

Date time	Profile	Region	Signals Involved
2024-04-05 01:51:16	20mm	CVR_L1,CVR_L2,PINCHROLL_L1,PINCHROLL_L2,STAND_7_12,STAND_13_18,CVAH_L2	[12:23]_FFB Line 2 RPM Act, [12:24]_FFB Line 2 Line Speed, [17:50]_TB2B Motor Current, [5:44]_L2_DCVRD Actual Position for PID
2024-04-15 13:25:09	12mm	CVAH_L1, PINCHROLL_L2	[1:59]_E35_CVAH_Actual Torque, [1:33]_L1_PR4_ActTorque, [5:19]_L2_PR2_ActSpeed and [5:22]_L2_PR3_ActSpeed

Table 6.1.4.2.5: False Positive Case Studies Summary

6.1.4 Optimizing Grafana Dashboard Configuration for Import

Sharing Grafana dashboards as JSON objects for import into other systems can be hindered by configuration replacements during the import process. Previously, modifying bucket names for individual panels required a cumbersome manual approach after configuring the InfluxDB data source: locating the panel, accessing its query settings, and manually updating the bucket name within the InfluxDB query. Additionally, some systems might necessitate adjustments to the UID and type within the dashboard's JSON model (updating UID and changing type from "influxdb" to the new bucket name). To address these inefficiencies and reduce manual errors, a Python script was developed. This script streamlines the process by reading and parsing the dashboard JSON file, identifying and replacing designated configurations (bucket name, UID, and type), and offering flexibility to either generate a new JSON file with the updates or directly modify the existing one.

6.2 Interpretation of Results

In this section, we will discuss the interpretation of the key outcomes and their broader significance within the context of the system's performance and operational efficiency.

6.2.1 Binary Signals Visualization

Significance:

Visualizing binary signals through bar charts significantly enhances their interpretability. This approach allows for efficient and accurate analysis of signal state durations, which can be crucial for understanding system behavior.

Example:

Consider an example of a binary HMD (name of a region) signal with a series of 0s and 1s, where the width of each state is important. Imagine a binary signal that switches between 0 and 1 as follows:

0-0-1-1-1-0-1-1

Here, the width of the '1' states (three 1s in the middle and two 1s at the end) is crucial for determining the duration of specific conditions.

Bar Chart:

To facilitate a direct comparison of signal state durations, the corresponding bar chart will be displayed. Converting the above signal to a bar chart would result in bars of varying lengths representing the continuous '0' and '1' states. For instance, the three continuous 1s would be shown as a single bar three units long, and the two continuous 1s would be shown as a single bar two units long. This visual format makes it immediately apparent which states are longer and by how much, facilitating quick and accurate comparisons.

Visualization Format:

The figure presents a two-part visualization of the binary signals.

Fig 6.2.1.1: Binary Signals Original and Bar Graph Plotting

- **Top Panels:** Display the original plots of the signals. These plots allow us to directly observe the transitions between the two states (0 and 1) over time.
- **Bottom Panels:** Depict the distribution of signal widths for both states (0 and 1). This visualization is typically presented as a bar chart, where the x-axis represents the duration (width) of each state and the y-axis represents the frequency of occurrence for each duration.

To enhance visual interpretation, the bars within the chart are typically color-coded. Red bars commonly represent the width of the "1" state, while green bars represent the width of the "0" state. This color scheme facilitates the quick identification and comparison of the duration for each signal state.

6.2.2 Motif Integration for Anomaly Detection

Significance:

Incorporating motifs into the dashboard provides a valuable tool for anomaly detection. Motifs capture the essence of the signal's behavior in a concise, generalized pattern. This allows for efficient and accurate identification of deviations from the expected behavior, aiding in anomaly detection.

Motif vs. Raw Signal:

Unlike directly observing the raw signal pattern, which can be challenging for comparison, motifs present a readily understandable and repeatable pattern. This motif is displayed alongside the original signal plot on the dashboard.

Anomaly Detection with Motifs:

By visually comparing the original signal and the motif, users can readily identify deviations from the expected pattern. This allows for a more efficient and accurate assessment of whether the signal behavior is normal or anomalous.

6.2.3 Dynamic Discord Highlighting with Python Script

Significance:

The Python script employing dynamic discord detection offers several key advantages over traditional anomaly detection methods:

1. **Real-time Discord Detection:** The Python script facilitates real-time calculation of discords, ensuring that the discord detection process remains up-to-date with the latest data.
2. **Flexibility and Customization:** Python provides a powerful and flexible environment for implementing various anomaly (discord) detection algorithms.
3. **Enhanced Visualization in Grafana:** The dynamic tagging of anomalies within the Python script enables clear visual representation of anomalies within the dashboard.

6.2.4 Case Studies

Significance of Case Studies in Algorithm Analysis and Validation

Case studies hold immense significance in the realm of algorithm analysis and validation. They provide a crucial opportunity to examine the practical performance of an algorithm within a specific context. Dashboards, like the one developed in this project, serve as invaluable tools for case study construction. By visualizing algorithm outputs and system behavior, dashboards enable researchers to gain a deeper understanding of the algorithm's effectiveness.

The validation process within a Business Support Platform (BSP) serves a vital role in identifying areas for improvement. By comparing the algorithm's outputs with real-world data in the BSP, case studies can reveal instances where the algorithm:

- **Generates True Positives:** This occurs when the algorithm correctly identifies anomalies or patterns of interest within the data. These successes validate the algorithm's ability to function as intended.
- **Produces False Positives:** This happens when the algorithm mistakenly identifies normal behavior as anomalous. Case studies uncover these false alarms, allowing for refinement of the algorithm to reduce such errors.

Through a combination of dashboard visualizations and BSP validation, case studies offer a comprehensive evaluation of an algorithm's strengths and weaknesses. This information is essential for fine-tuning the algorithm and ensuring its robustness in real-world applications.

6.3 Comparison with Objectives

All objectives of the project were achieved. Enhanced dashboards within Grafana played a vital role in communicating the project's findings and facilitating real-time process monitoring. Key enhancements incorporated into the dashboards included improved binary signal visualization, where binary signals were readily identifiable based on their period (".") notation, allowing for clear distinction between active and inactive states of specific equipment. Bar charts were implemented to effectively showcase the durations of these binary signals, enabling personnel to easily compare signal durations and identify potential process variations. Static motif images were incorporated as reference points for anomaly detection, serving as independent time series visualizations that provided a visual representation of the established normal patterns for specific sensors.

Additionally, a Python script was developed to dynamically calculate anomalies ("discords") for each data sequence within Grafana, facilitating the automatic highlighting of discord events on the dashboards and ensuring their easy identification by plant personnel. These enhancements transformed the dashboards into informative and user-friendly tools for real-time process monitoring and anomaly detection at the BSP mill.

6.4 Discussion of Key Findings

Enhanced dashboards within Grafana played a vital role in communicating the project's findings and facilitating real-time process monitoring. Key enhancements incorporated into the dashboards included:

- Improved Binary Signal Visualization: Binary signals were readily identifiable based on their period (".") notation, allowing for clear distinction between active and inactive states of specific equipment.
- Bar Charts for Signal Durations: Bar charts were implemented to effectively showcase the durations of these binary signals. This visual representation enabled personnel to easily compare signal durations and identify potential process variations.

6.5 Limitations and Future Directions:

Case studies falter for irregular signals. Comparing unpredictable patterns, potentially missing anomalies, and needing multiple motifs from different times for full understanding hinder their effectiveness.

- **Difficulty in Applying to Irregular Patterns:** Case studies are most effective when analyzing signals with consistent or predictable patterns. When dealing with irregular signals, the very essence of a case study - a focused examination of a particular instance - becomes problematic. Imagine trying to understand a language by studying just one, unusual sentence. Irregular signals require a broader approach that can capture the full range of variability.
- **Challenges in Motif Comparison:** Motifs, which are recurring patterns within a signal, are often used for comparison in case studies. However, for irregular signals, comparing motifs becomes difficult. How similar do motifs need to be to be considered the "same"? What if there are significant variations within a single motif across different instances?
- **Discordant Data and False Positives:** Discord discovery approaches, which aim to identify deviations from expected patterns (discords), might not be perfect. They might miss some true discords (false negatives) or mistakenly identify regular variations as discords (false positives). Imagine a heart rate monitor flagging every slight beat variation as an abnormality. In the context of irregular signals, where some level of variation is expected, these approaches need careful calibration to avoid misleading results.
- **Need for Multiple Motifs:** In some cases, a single motif might not be sufficient to capture the complexity of an irregular signal. For example, analyzing a patient's sleep patterns might require considering separate motifs for deep sleep, REM sleep, and wakefulness. This suggests the need for incorporating multiple motifs from different timeframes or contexts within the same profile for a more comprehensive understanding.

By acknowledging these limitations, researchers can develop more robust methods for analyzing irregular signals. This might involve using alternative comparison techniques, incorporating statistical analysis alongside case studies, or employing machine learning algorithms that can learn from larger datasets encompassing a wider range of variability.

CHAPTER – VII

LEARNING OUTCOME

CHAPTER VII

LEARNING OUTCOME

This section summarizes the key takeaways from the internship project, reflects on its achievements, and explores potential avenues for future development.

7.1 Summary of Findings

The internship focused on enhancing Grafana dashboards for improved system monitoring and anomaly detection. Key findings include:

1. Enhanced Interpretability with Bar Charts for Binary Signals:

- **Problem:** Traditional line graphs for binary signals (on/off) made it difficult to gauge the duration each state persisted. This could lead to delays in identifying potential issues.
- **Solution:** Implemented bar charts specifically for binary signals. These charts clearly show the duration each state (on/off) lasted, allowing for faster analysis and easier identification of prolonged states that might indicate problems.
- **Impact:**
    - o Reduced time spent interpreting signal behavior.
    - o Improved ability to spot potential issues quickly based on state durations.

2. Motif-based Anomaly Detection for Efficient Identification:

- **Problem:** Traditional anomaly detection methods in Grafana dashboards might be generic and miss specific behavioral patterns.
- **Solution:** Introduced motifs, which are recurring patterns of signal behavior. By integrating motif identification into the dashboard, users can easily see when the signal deviates from expected patterns, enabling efficient anomaly detection.
- **Impact:**
    - o More targeted anomaly detection focused on significant deviations from normal behavior.
    - o Reduced time spent analyzing generic anomaly alerts.

3. Real-time Discord Highlighting with Python Script:

- **Problem:** Existing anomaly detection methods might not provide real-time identification or clear visual representation within the dashboard.
- **Solution:** Developed a Python script that leverages dynamic discord detection. This script continuously analyzes data and highlights anomalies in real-time within the dashboard, enhancing situational awareness and enabling rapid response.
- **Impact:**
    - Faster identification of anomalies as they occur.
    - Clear visual representation of anomalies within the dashboard for immediate attention.

4. Streamlined Configuration Management with Python Script:

- **Problem:** Large-scale deployments with numerous Grafana dashboards could lead to time-consuming and error-prone configuration management during JSON import.
- **Solution:** Created a Python script that streamlines the configuration process. This script automates tasks during JSON import, significantly reducing configuration time and minimizing errors, especially for large deployments.
- **Impact:**
    - Saved significant time during dashboard configuration.
    - Reduced the risk of errors during configuration management.

These findings demonstrate the effectiveness of the implemented enhancements in achieving the project's objectives of improved data visualization, efficient anomaly detection, and streamlined dashboard management. This dashboard upgrade enhanced user experience in several ways. Visualizations made signal state durations easier to analyze, motif integration aided anomaly detection, and a Python script provided real-time highlighting of anomalies.

7.2 Achievement of Objectives

The project successfully achieved its primary objectives:

- **Improved Data Visualization:** The visualization techniques implemented (bar charts and motifs) demonstrably enhanced the clarity and interpretability of data within the Grafana dashboard.

- **Enhanced Anomaly Detection:** The integration of motifs and the dynamic discord highlighting script significantly improved the efficiency and accuracy of anomaly detection capabilities.
- **Streamlined Workflow:** The Python script for automated configuration management streamlined the process of importing Grafana dashboards, promoting efficiency and minimizing human error.

Challenges encountered during the project included:

- **Complexity of Anomaly Detection Algorithms:** Selecting and implementing appropriate anomaly detection algorithms within the Python script required careful consideration to ensure accuracy and avoid computational overhead.
- **Fine-Tuning Visualization Techniques:** Striking a balance between detailed information and visual clutter in the dashboard required ongoing refinement of visualization techniques.

These challenges were addressed through ongoing research, experimentation, and collaboration with project mentors.

7.3 Implications and Recommendations

The project findings have several key implications:

- **Improved System Monitoring:** Enhanced data visualization and anomaly detection capabilities contribute to improved system monitoring, allowing for faster identification and resolution of potential issues.
- **Data-driven Decision Making:** The project's advancements enable users to make more informed decisions based on visualized data and real-time anomaly alerts.
- **Scalability and Efficiency:** The automated configuration script promotes scalability and efficiency, particularly for managing numerous dashboards or complex configurations.

Based on these findings, recommendations for future developments include:

- **Exploration of Advanced Anomaly Detection Techniques:** Implementing more sophisticated anomaly detection algorithms within the Python script could further enhance the accuracy and flexibility of the system.

- **Integration with Machine Learning Models:** Exploring the integration of machine learning models for anomaly detection could potentially improve the system's ability to identify complex and evolving patterns.
- **User Interface Refinement:** Continuously refining the user interface of the dashboard can enhance user experience and facilitate efficient data exploration.

7.4 Future Scope

The project opens doors for exciting future research and development avenues:

- **Investigation of Domain-Specific Anomaly Detection:** Tailoring anomaly detection algorithms to specific domains (e.g., network traffic analysis) could further enhance the effectiveness of the system.
- **Self-Learning Anomaly Detection:** Exploring self-learning algorithms within the Python script could enable the system to adapt and improve anomaly detection capabilities over time.
- **Cloud-based Deployment:** Deploying the enhanced dashboard solution on a cloud platform would promote wider accessibility and scalability.

These avenues present significant opportunities to further refine and extend the project's functionalities.

7.5 Personal Reflections

My internship experience proved to be a transformative journey that significantly accelerated my professional growth. Here's a closer look at some key takeaways:

- **Technical Expertise:** I went beyond theoretical knowledge, gaining hands-on experience in data visualization. I explored various techniques, like crafting informative bar charts for binary signals, which improved my ability to translate data into clear and actionable insights. Additionally, I delved into anomaly detection, learning to identify and highlight deviations from expected patterns. This equips me to proactively address potential issues before they escalate. Furthermore, I honed my Python scripting skills, developing a script that automates configuration tasks, saving valuable time and minimizing errors, especially for large deployments.

- **Sharpened Problem-Solving:** Throughout the project, I encountered real-world challenges that demanded creative solutions. This fostered my problem-solving abilities. I learned to effectively break down complex issues, research potential solutions, and ultimately implement effective strategies. This newfound skillset will be invaluable in tackling future obstacles in data science.
- **Enhanced Communication and Collaboration:** The internship provided a platform to develop my communication and collaboration skills. Working alongside mentors and colleagues, I learned to articulate technical concepts clearly and concisely. I actively participated in discussions, seeking and incorporating valuable feedback. This collaborative environment fostered a sense of teamwork and instilled the importance of effective communication in achieving shared goals.

In conclusion, this internship wasn't just about acquiring technical skills; it significantly contributed to my overall academic and career development. The practical experience solidified my passion for data science. Witnessing the impact of data-driven solutions firsthand has motivated me to further explore and refine my skills within this dynamic field. This internship has equipped me with the technical expertise, problem-solving abilities, and collaborative spirit to thrive in the exciting world of data science.

It has solidified my passion for Data Science and motivated me to pursue further exploration and learning in this domain.

CHAPTER VIII

CONCLUSION AND FUTURE SCOPE

CHAPTER VIII

CONCLUSION AND FUTURE SCOPE

This internship provided a valuable learning experience that equipped me with new skills, knowledge, and insights that will propel me forward in my academic and professional journey.

8.1 Skills Developed

Throughout the internship, I had the opportunity to develop and enhance several key skills:

8.1.1 Technical Skills

My internship transformed my technical skillset, particularly in the realm of data communication and analysis. I dove deep into data visualization, mastering tools like Grafana. This newfound proficiency empowered me to craft dynamic and insightful visual representations of complex datasets. No longer just interpreting data, I could now translate it into clear, compelling narratives, effectively communicating critical insights to stakeholders. This ability to transform data into actionable knowledge proved invaluable.

Furthermore, I ventured into the fascinating world of anomaly detection, a cornerstone of data analysis. Through extensive Python scripting, I embarked on a journey to explore various anomaly detection algorithms. This hands-on experience wasn't just about memorizing formulas; it was about truly understanding the theoretical underpinnings of these algorithms and their practical applications.

I meticulously developed, tested, and refined models capable of identifying outliers and unusual patterns within data. This newfound expertise is crucial for maintaining data integrity and driving informed decision-making – a skill set coveted by any organization.

Finally, the internship wasn't just about these specialized skills; it also provided a solid foundation in essential tools. I honed my proficiency in Python, a powerful programming language that empowers data manipulation, automation of tasks, and the implementation of various data analysis techniques. Additionally, I solidified my grasp of Microsoft Office, an indispensable suite for any professional setting. This well-rounded skillset equips me to tackle data challenges with confidence, extract meaningful insights, and effectively communicate them to a wider audience.

8.1.2 Problem-Solving

The project I worked on posed numerous challenges that required creative and systematic problem-solving approaches. Faced with intricate issues, I learned to break them down into manageable components, identify potential obstacles, and devise strategic solutions. This process involved extensive research to explore possible methodologies, testing different approaches, and iterating on solutions to optimize outcomes.

By systematically tackling each problem, I not only resolved immediate project-related issues but also developed a robust problem-solving framework that I can apply to future endeavors. This iterative approach reinforced my ability to think critically and adapt to evolving project needs, ensuring efficient and effective solutions.

8.1.3 Problem-Solving

Effective communication and collaboration were pivotal to my success during the internship. Engaging with mentors and colleagues provided a rich environment for developing these skills. I learned to clearly articulate technical concepts, ensuring that my explanations were accessible to both technical and non-technical audiences. This clarity in communication facilitated smoother project progress and fostered a collaborative atmosphere.

Active listening was another critical component of my development. By attentively considering feedback from mentors and peers, I was able to refine my ideas and approaches, leading to more robust outcomes. Working within a team environment also highlighted the importance of collaboration, as I contributed to group discussions, shared responsibilities, and supported collective goals. This collaborative experience underscored the value of diverse perspectives and teamwork in achieving project success.

8.2 Knowledge Gained

This internship deepened my understanding of:

- **Data Visualization:** During my internship, I gained practical experience in using data visualization techniques to enhance the interpretability of complex data for improved system monitoring. This involved mastering tools like Grafana to create dynamic and interactive dashboards that effectively communicate trends, anomalies, and real-time system health. This proficiency in data visualization allows me to translate complex datasets into clear, actionable insights for stakeholders.

- **Anomaly Detection:** I further expanded my skillset by exploring various anomaly detection algorithms and their application in real-world scenarios. Through extensive Python scripting, I gained a deep understanding of the theoretical underpinnings of these algorithms, allowing me to develop, test, and refine models capable of identifying outliers and unusual patterns within data. This expertise in anomaly detection is crucial for maintaining data integrity and proactively identifying potential system issues.

- **Workflow Optimization:** Recognizing the importance of efficiency, I actively sought opportunities to optimize workflows. Notably, I learned to automate tasks like Grafana dashboard configuration using Python scripting. This experience highlighted the significant efficiency gains achievable through automation, allowing me to focus on more strategic data analysis tasks while minimizing the risk of human error.

8.3 Professional Development

My internship experience served as a significant catalyst for my professional development, solidifying my passion for Data Science and shaping my long-term career aspirations. This hands-on immersion not only equipped me with valuable technical skills but also fostered a deeper understanding of the practical applications of data analysis in real-world scenarios. Additionally, it enhanced my problem-solving abilities and boosted my confidence in handling complex data-driven projects.

- **Igniting a Data Science Passion:** While my interest in Data Science may have been theoretical before the internship, this experience provided a real-world crucible. Here, I actively applied my knowledge in a practical setting. Witnessing the power of data visualization and anomaly detection in unlocking insights and optimizing systems firsthand ignited a fervent passion within me. This internship was instrumental in transforming a nascent interest into a resolute career ambition.
- **Unearthing Strengths and Refining Weaknesses:** Self-awareness is paramount for professional growth, and the internship provided a valuable opportunity for self-reflection. It not only highlighted my strengths but also areas for improvement. I discovered a natural aptitude for acquiring new technical skills, particularly in data visualization tools like Grafana and anomaly detection algorithms implemented through Python scripting. However, I also identified the need to further develop my communication skills, specifically when presenting highly technical concepts to diverse audiences. By acknowledging these strengths and weaknesses, I am empowered to create a targeted development plan to solidify my skillset and become a well-rounded data scientist.
- **Crystallizing Career Goals:** Career aspirations can often be broad and lack a clear direction. This internship experience served as a focusing lens, clarifying my career path. I am now more driven than ever to pursue a role that allows me to leverage my technical expertise in data analysis and system monitoring. Additionally, the collaborative nature of the internship environment sparked a passion for working within a team to develop innovative solutions. My ideal career path now encompasses not just technical proficiency but also the ability to collaborate effectively and translate complex concepts into clear communication for stakeholders.

8.4 Personal Growth

Engaging with colleagues, clients, and stakeholders improved my ability to convey ideas clearly and confidently. Beyond technical skills and professional development, the internship fostered significant personal growth:

- **Challenge Overcoming:** The internship wasn't without its challenges. A particularly rewarding one involved selecting appropriate anomaly detection algorithms for the project. Different algorithms excel at identifying specific types of anomalies, making the choice crucial. I embarked on extensive research, delving into the strengths and weaknesses of various algorithms offered by Python libraries like scikit-learn. Collaborating closely with senior data scientists, I actively participated in discussions, asking insightful questions to understand the underlying data characteristics and potential anomalies we aimed to detect. This collaborative approach, coupled with my research, allowed us to identify the optimal algorithms for our specific needs.

- **Lessons Learned:** The internship instilled several invaluable lessons that will undoubtedly benefit my future career. First and foremost, it emphasized the importance of continuous learning. The field of Data Science is constantly evolving, with new techniques and technologies emerging frequently. This experience ignited a passion for staying current; I actively seek out new resources, attend webinars, and participate in online communities to stay at the forefront of the field. Furthermore, the internship highlighted the importance of adaptability and a willingness to experiment. Not every approach will succeed in the first attempt. I learned the value of analyzing failures, adjusting strategies, and experimenting with different techniques – a crucial skill for problem-solving success.

- **Perspective and Values:** The collaborative environment of the internship fostered not only my technical skills but also my professional values. Working alongside a diverse team of data scientists allowed me to appreciate the value of different perspectives. Each team member brought a unique set of experiences and knowledge to the table. For example, a colleague with a strong background in statistics might have focused on identifying statistically significant deviations in the data, while another with experience in machine learning might have explored anomaly detection algorithms.

8.5 Future Applications

Empowered by the internship's data visualization and anomaly detection expertise, I plan to propel my Data Science career, continuously learning new techniques, and applying the problem-solving approach of research, experimentation, and collaboration to future challenges.

- **Academic Studies:** The insights gained from data visualization and anomaly detection will be valuable for future academic projects and research activities. This skill is crucial for academic research, as it aids in the clear presentation of findings, making it easier to communicate results effectively to peers, professors, and the broader academic community. Furthermore, anomaly detection has taught me how to identify and address outliers and irregularities in datasets, which is essential for ensuring the accuracy and reliability of research outcomes.

- **Future Career:** My newly acquired technical skills will be directly applicable to future job opportunities in Data Science. Furthermore, the problem-solving and communication skills honed during the internship will be valuable assets in any professional setting. This practical knowledge equips me with the tools necessary to tackle real-world challenges and deliver data-driven solutions in a professional setting. Additionally, the problem-solving skills I developed—learning to approach complex problems methodically and finding innovative solutions—are invaluable in any career.

- **Lifelong Learning:** The internship instilled in me the importance of continuous learning and adaptation. I plan to actively seek opportunities to learn new skills and stay updated with advancements in Data Science. The internship experience emphasized the need for a proactive approach to learning, whether through formal education, professional development courses, or self-study.

This internship provided a transformative learning experience, equipping me with advanced technical skills in data visualization and anomaly detection, and reinforcing my problem-solving and communication abilities. My mastery of tools like Grafana and Python, alongside a deep understanding of anomaly detection algorithms, has prepared me to tackle complex data challenges. The experience has been pivotal in my academic pursuits, offering valuable insights for research, and has solidified my passion for a Data Science career by crystallizing my career goals and highlighting the importance of continuous learning and adaptability.

REFERENCES

- DTAIDistance Developers. (n.d.). https://dtaidistance.readthedocs.io/en/latest/modules/dtw.html
- Grafana Labs. (n.d.). *Grafana Documentation*. https://grafana.com/docs
- InfluxData, Inc. (n.d.). *InfluxDB Documentation*. https://docs.influxdata.com/influxdb/latest/
- Mueen, A., Keogh, E., & Batista, G. (2006, December). *Finding Temporal Motifs in Time Series Data*. In Proceedings of the Sixth IEEE International Conference on Data Mining (ICDM'06) (pp. 796-801). https://www.cs.ucr.edu/~eamonn/Motif_Discovery_ICDM.pdf
- Stumpy Documentation. (n.d.). https://stumpy.readthedocs.io/en/latest/Tutorial_STUMPY_Basics.html.

ANNEXURE – I

भारतीय प्रौद्योगिकी संस्थान भिलाई
जिला-दुर्ग, छत्तीसगढ़-491001
Indian Institute of Technology, Bhilai
Dist.- Durg, Chhattisgarh - 491001
Website: www.iitbhilai.ac.in

No.: IITBH/R&D/2024-25/144 Date: June 19, 2024

TO WHOM IT MAY CONCERN,

This is to certify that Ms. Madhurima Rawat was working as an Intern at IIT Bhilai from 23rd Feb 2024 to 12th June 2024 in the project "DIGITAL TRANSFORMATION SYSTEM FOR PRE-FAILURE ALERT GENERATION FOR EQUIPMENT FAILURE & COBBLE REDUCTION BASED ON DATA ANALYTICS AND VIDEO ANALYTICS AT BRM" at IIT Bhilai under project investigator (PI) Dr. Gagan Raj Gupta. During this period, she worked on on utilizing Python for motif and discord analysis within the project. Additionally, she contributed to enhancing the Grafana dashboard by:

- Integrating motifs (normal patterns) for visualization.
- Implementing dynamic discord detection based on data from the BSP mill server.
- Creating bar graphs to visualize binary signals within the dashboard.

She has also participated in case studies based on anomaly events identified from the BSP mill server data.
We wish her all the best and success in her future endeavors.

Deputy Registrar
Research & Development