INDUSTRIAL TRAINING REPORT

ON

PREDICTIVE MODELLING OF SCALING PHENOMENA USING INDUSTRIAL SENSOR DATA

AT

Kuantum Papers Mill, Saila Khurd, Distt. Hoshiarpur, Punjab 144529

Submitted By:

Aditi Mahajan (23106007)

Vani (23106098)

Supervised By:

Mr. Ashwani Kumar

Sr. DGM (TECH EXCELLENCE)

TRAINING PERIOD (16 JUNE 2025 TO 25 JULY 2025)

Dr. B. R. Ambedkar National Institute of Technology, Jalandhar Department of Instrumentation and Control Engineering

Instrumentation and Control Engineering

CANDIDATE'S DECLARATION

I hereby certify that the work presented in the report entitled "Predictive Modelling of Scaling Phenomena Using Industrial Sensor Data" is prepared by me during a period from 16th June 2025 to 25th July 2025 under the supervision of Mr. Ashwani Kumar , SR. DGM(Tech Excellence) at KUANTUM PAPERS Ltd.

Dated: 25-07-2025

Aditi Mahajan
(23106007)

SUPERVISOR'S DECLARATION

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dated: 25-07-2025

Ashwani Kumar
Supervisor

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to **Mr. Nandivardhan Morareddy**, **Mr. Ashwani Sharma**, **Mr. Rajeev Mahajan, Mr. Kamaljit Kamboj,** and **Ms. Varsha Singh** of Kuantum Paper Mill for their invaluable support and guidance throughout the course of my internship. Their mentorship played a crucial role in enhancing my understanding of real-time industrial processes and expanding my technical knowledge in the instrumentation domain. I would also like to thank **Mr. Ajay Sharma (HR)** for giving me this opportunity.

Their constant support and guidance enabled me to first gain a comprehensive understanding of the various processes involved in the pulp and paper mill, particularly focusing on the behavior of process variables during production. This foundational knowledge helped me identify critical parameters that influence scaling in equipment. Building on this, I analyzed real-time sensor data using machine learning techniques to develop a predictive model capable of detecting scaling conditions. This project not only strengthened my grasp of industrial operations but also equipped me with practical skills in data analysis, model development, and predictive maintenance strategies.

It has been a privilege to learn under the supervision of such dedicated and knowledgeable professionals.

CONTENTS

INTRODUCTION TO KUANTUM PAPERS LIMITED

1.1 Introduction

Kuantum Papers Limited, established in 1980, located in Saila Khurd, Punjab, is a leading agro and wood-based paper manufacturing company in India. It specializes in producing high-quality writing and printing papers using eco-friendly raw materials like wheat straw, baggasse, and wood residues which would otherwise flare up and cause pollution. With a current production capacity of around 450 tonnes per day, the company is known for its wide range of products including Maplitho, Cream Wove, and various specialty papers. Kuantum is actively pursuing sustainability through social forestry, water recycling, and energy-efficient technologies, and is undergoing major expansions and digital transformations to increase efficiency and production capacity.

1.2 Vision & Mission

- To be the leading Paper Makers in an innovative and sustainable manner.
- Innovate continuously to enhance value in our business
- Pursue excellence in a sustainable manner
- Deliver exceptional value to all stakeholders

1.3 Manufacturing Plants

Kuantum Papers Limited operates a fully integrated paper manufacturing facility located in Saila Khurd, Hoshiarpur, Punjab. The plant specializes in the production of high-quality wood-free writing and printing papers, primarily using agro-residues such as wheat straw, bagasse, and other non-conventional fibers. With a current capacity of approximately 450 tonnes per day, the facility is equipped with advanced machinery including multiple paper machines, chemical recovery systems, pulping sections, and a co-generation power plant to ensure energy efficiency. In line with sustainable practices, Kuantum also employs a recovery boiler and modern effluent treatment plant to minimize its environmental footprint, showcasing its commitment to eco-friendly operations and circular resource use.

1.4 Products

Kuantum Papers Limited is engaged in the manufacturing and marketing of high-grade wood-free writing and printing papers made primarily from agro-residues and wood-based raw materials. The company produces a wide range of paper products under various GSM categories (ranging from 42 to 200 g/m^2), catering to diverse applications in the education, publishing, and corporate sectors. Key product categories include Maplitho, Cream Wove, and specialty papers such as bond, ledger, cartridge, cup-stock, coloured, and parchment papers. These products are known for their superior printability, smooth finish, and environmental sustainability. The products are widely used in the printing of books, trade directories, newsprint, diaries, calendars, and computer stationery.

1.5 Sustainability Initiatives:

Kuantum Papers Limited places a strong emphasis on sustainable and environmentally responsible manufacturing practices. The mill utilizes agro-residues like wheat straw and bagasse as core raw materials, thereby conserving forest resources and reducing the environmental impact associated with conventional wood-based papermaking. To further its sustainability goals, Kuantum operates a modern chemical recovery system, a high-efficiency recovery boiler, and an advanced effluent treatment plant, ensuring minimal environmental discharge and efficient energy utilization. The company also undertakes large-scale social forestry programs, encouraging local cultivation of pulpwood species and distributing saplings to farmers, which strengthens raw material availability and supports rural livelihoods. These initiatives reflect Kuantum's integrated approach toward eco-efficient operations and long-term environmental stewardship.

1.6 Operational and Financial Overview:

Kuantum Papers Limited delivered a steady financial performance in FY 2024–25, with a Total Operating Income (TOI) of ₹1,107.04 crore, reflecting an 8.8% decline compared to the previous year due to lower product pricing and rising input costs. The company reported a Profit After Tax (PAT) of ₹115.18 crore, down from ₹183.83 crore in FY 2023–24, primarily due to margin pressure. Despite the decline, Kuantum maintained a healthy Earnings Per Share (EPS) of ₹13.19 and demonstrated strong operational efficiency with a PBDIT of ₹242.59 crore. The company continues to invest in process optimization and sustainability, ensuring long-term financial stability while navigating market fluctuations.

2. Overview of the PAPER MILL

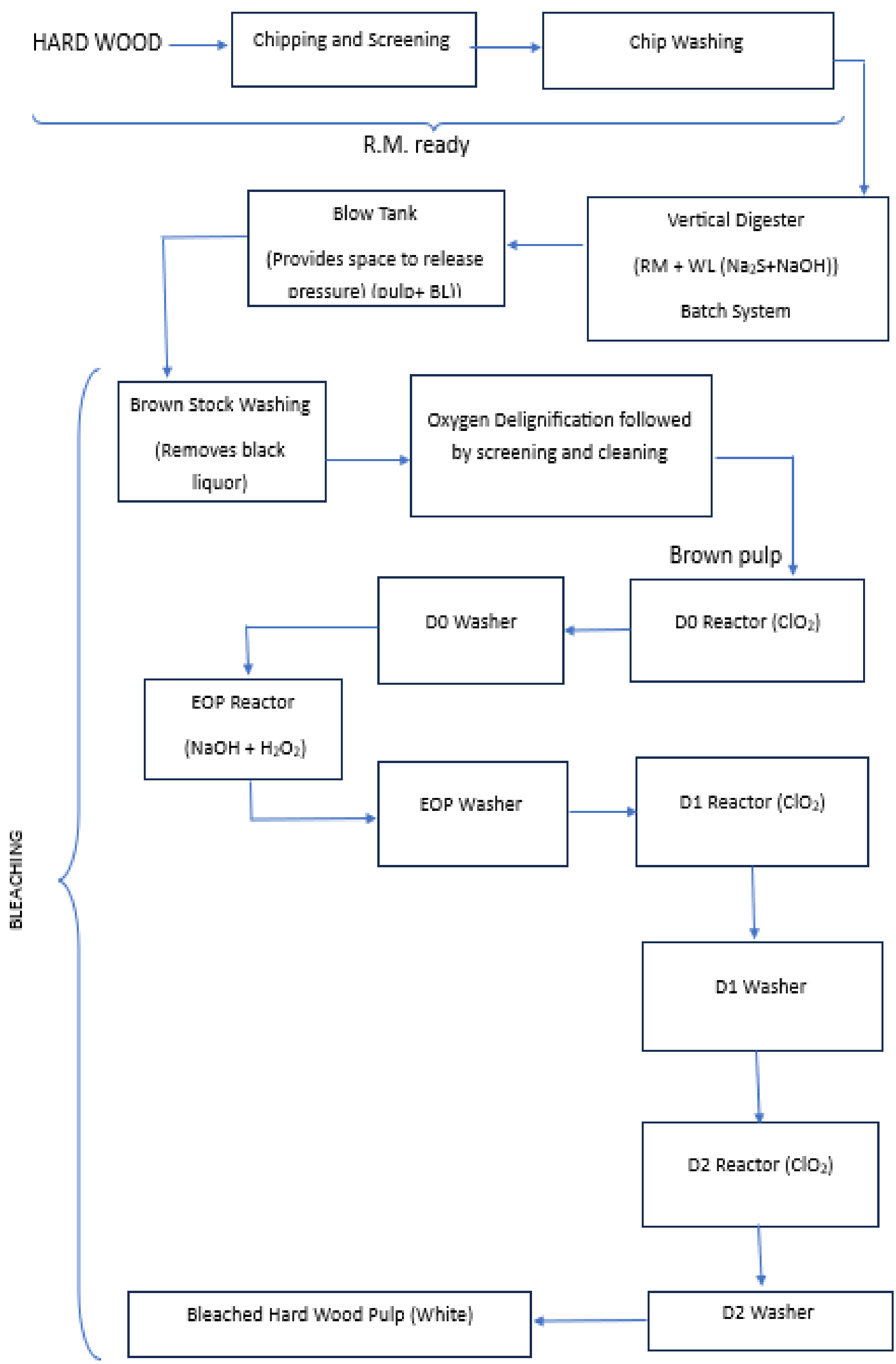
Kuantum Paper Mill is a prominent player in the pulp and paper industry, known for its environmentally conscious manufacturing and advanced industrial infrastructure. Situated in Punjab, India, the mill operates through a network of interdependent departments that ensure smooth and efficient production. The **Pulp Mill** converts agro-residues and wood-based raw materials into pulp, while the **Paper Machine Section** handles sheet formation, drying, and finishing. The **Chemical Recovery Plant** and **Effluent Treatment Plant (ETP)** manage resource recovery and environmental compliance. A dedicated **Power Plant** ensures uninterrupted energy supply for continuous operations. The **Instrumentation and Control Department** monitors and automates critical processes across the plant. Additionally, the **Technical Department** provides engineering support and process optimization, ensuring high operational standards. The **Human Resources (HR) Department** manages workforce development, training, and employee welfare, contributing to a professional and productive workplace. Together, these departments form a cohesive and efficient ecosystem driving Kuantum's success in the paper manufacturing sector.

Figure 1 Centri-cleaner in Pulp Mill

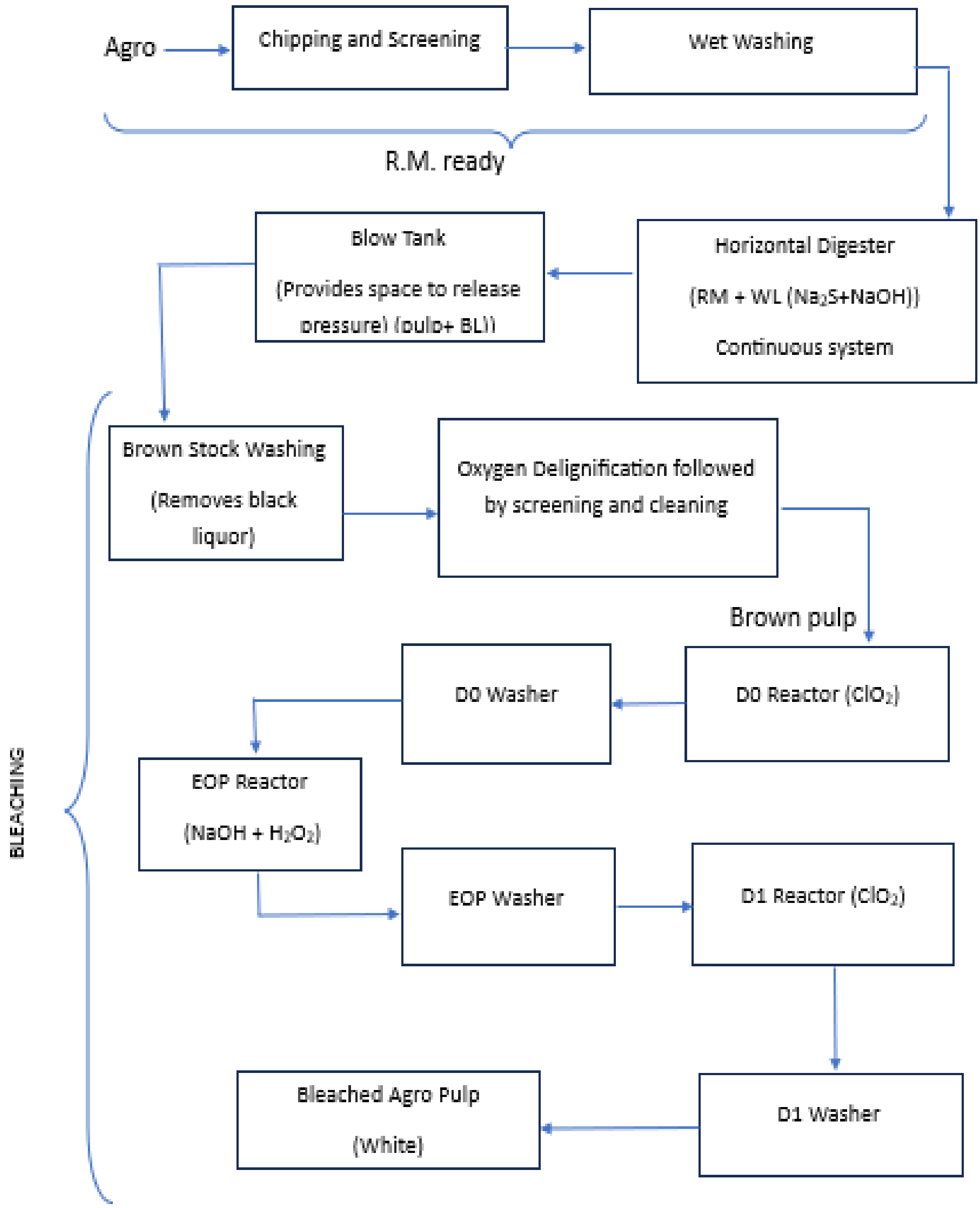
Figure 2 Paper Machine

Figure 3 Finishing House

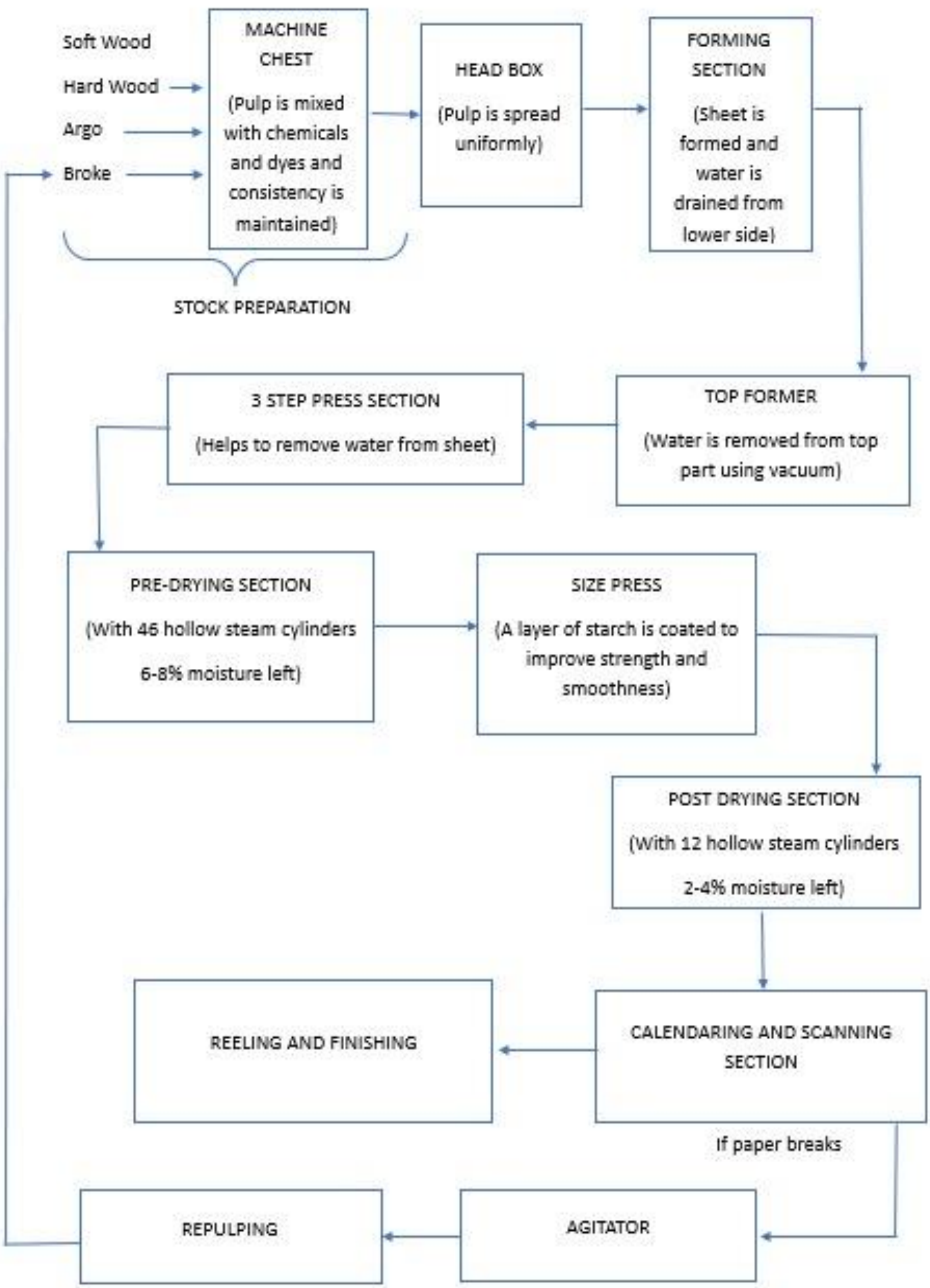
PULP MILL --- HARD WOOD



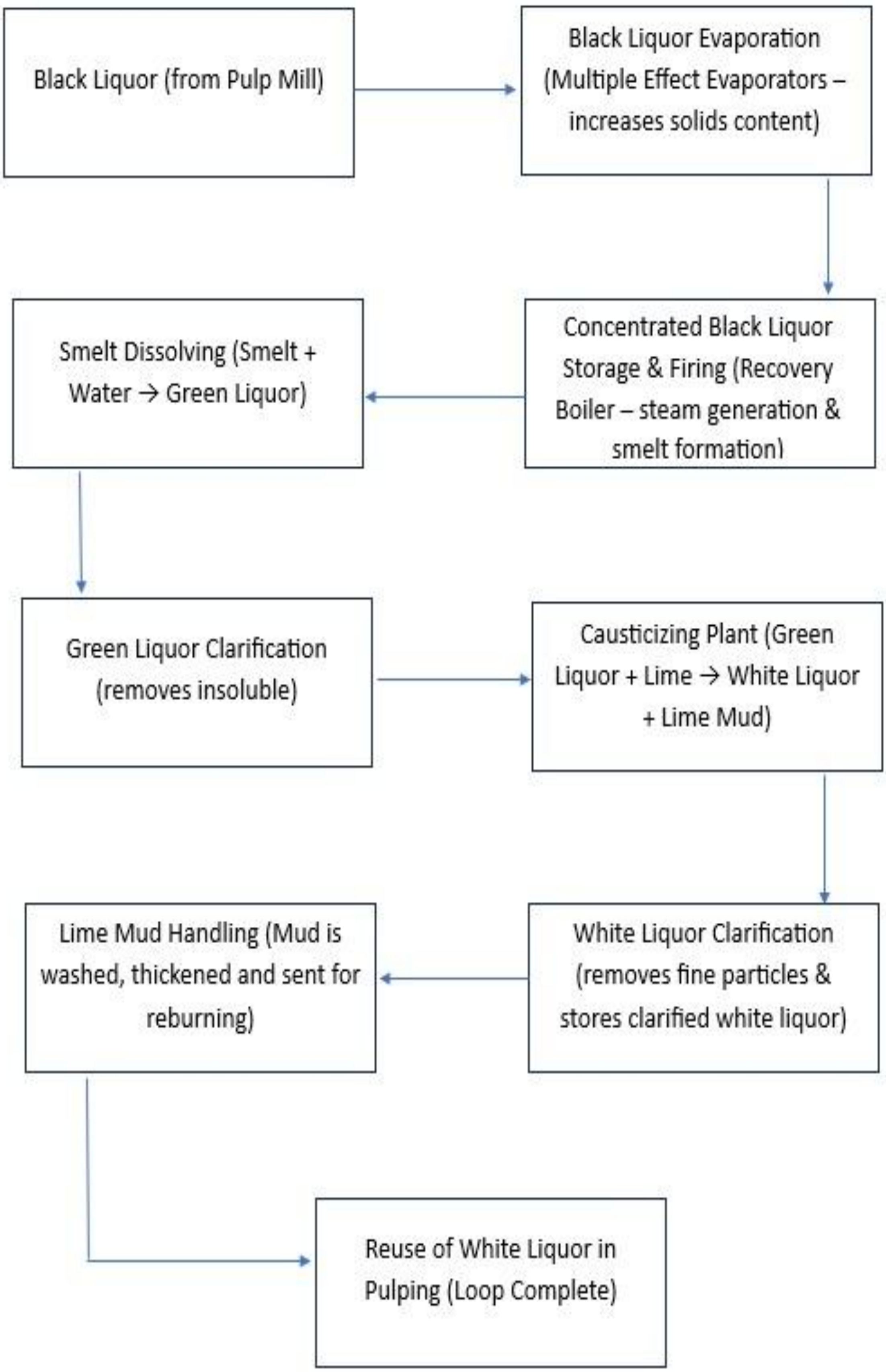
PULP MILL --- Agro



PAPER MACHINE

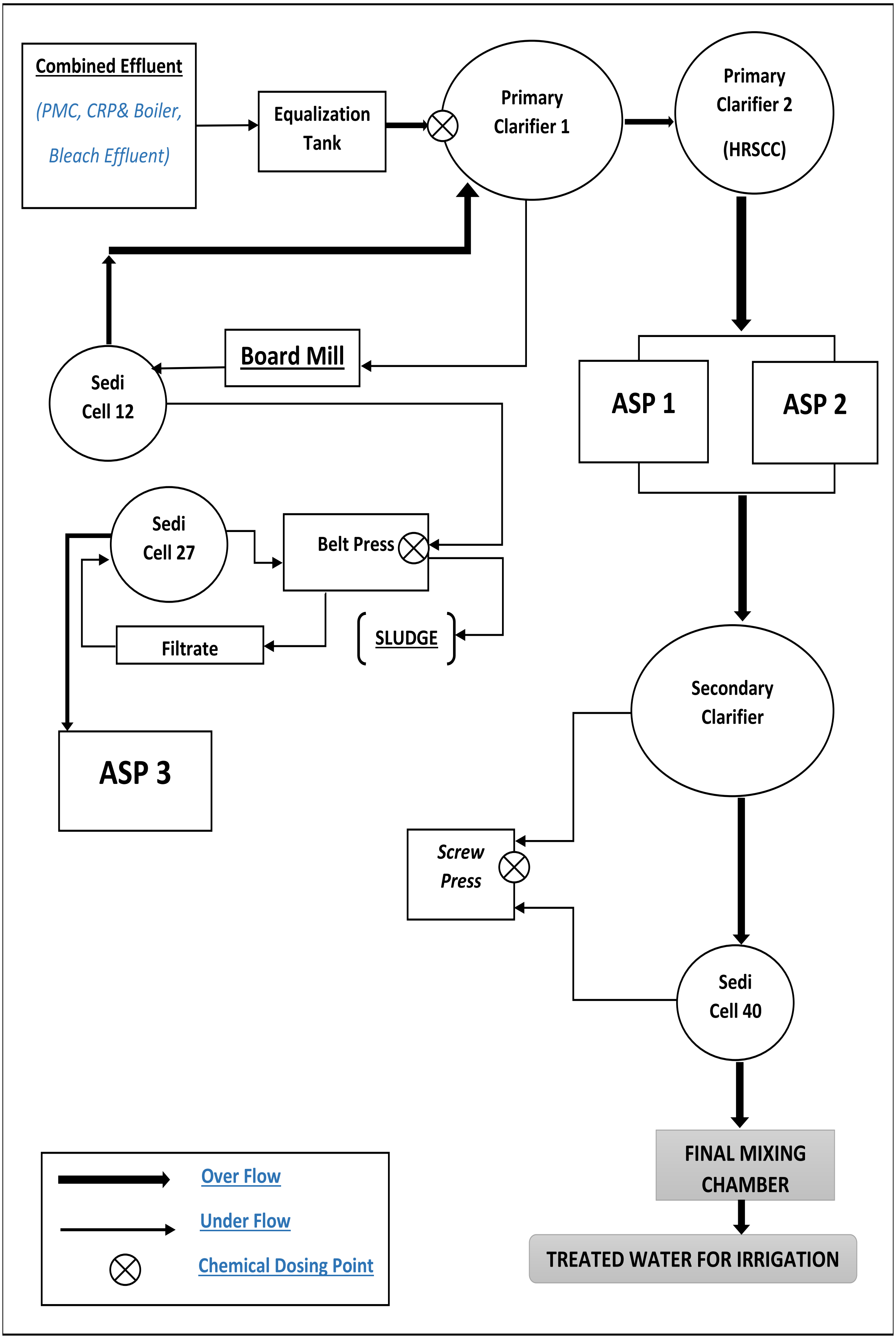


Chemical Recovery Plant

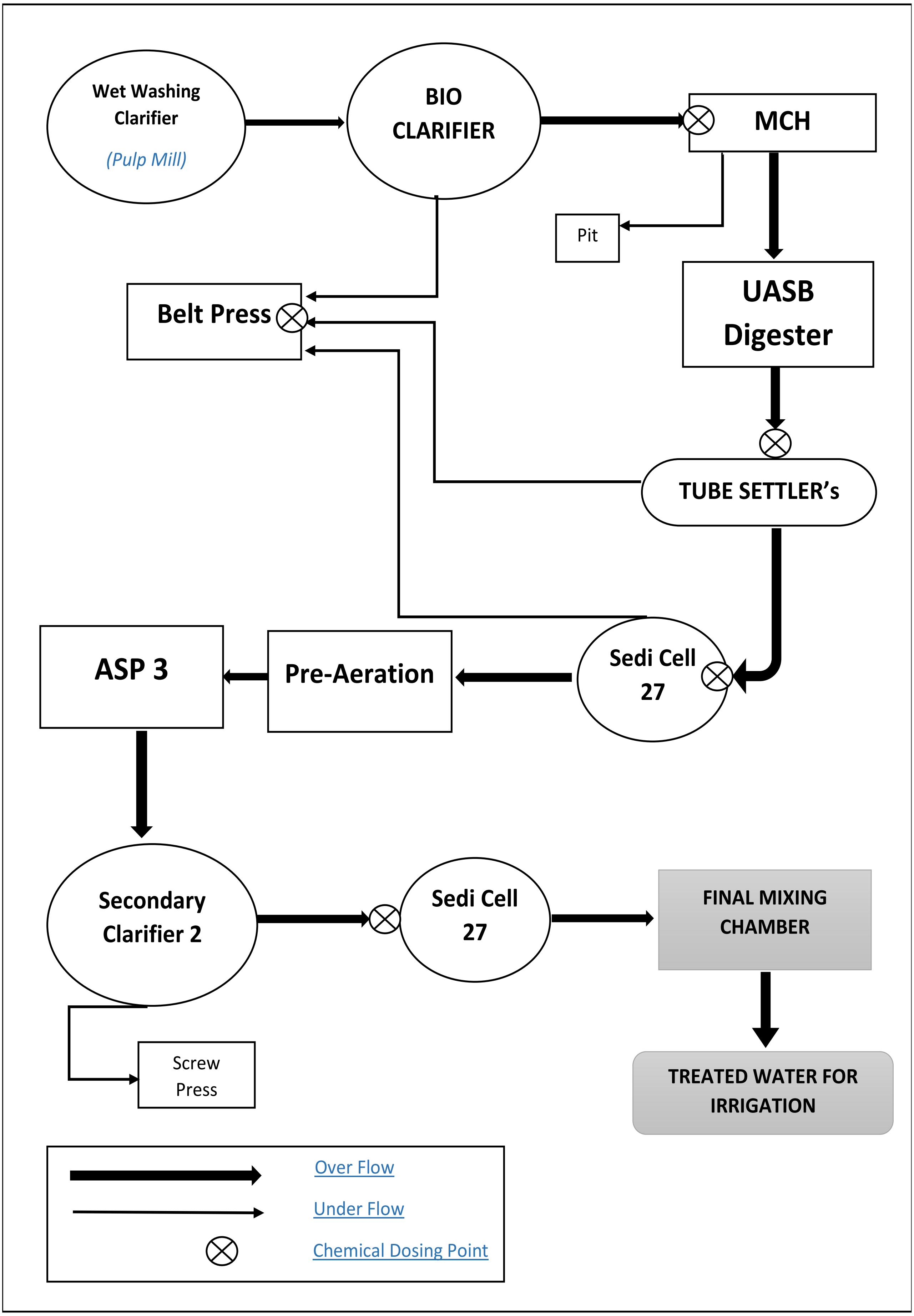


EFFLUENT TREATMENT PLANT

LOW COD STREAM



HIGH COD STREAM



3.On-Site Instrumentation: Devices and Applications

Instrumentation plays a critical role in ensuring the efficient, safe, and consistent operation of industrial processes in a paper mill. In **Kuantum Papers Limited**, where continuous monitoring and control of variables such as pressure, temperature, flow, and level are essential, field instrumentation serves as the backbone of process automation. Accurate and real-time data from instruments enable timely decision-making, reduce process variability, minimize downtime, and enhance overall product quality. From raw material handling to pulp preparation and final paper formation, every stage of production relies on precise instrumentation for smooth functioning and regulatory compliance. The integration of sensors, transmitters, controllers, and control systems like PLC and DCS ensures optimized performance and energy efficiency across the plant. Workshop is the important area of industry. In the workshop all engineering designs are implemented practically. Repair, maintenance and design work are done here.

Following are the various instruments and technologies seen and learnt by us:-

3.1 FLOW MEASUREMENT INSTRUMENTS :

Flow measurement plays a critical role in the process automation and control of modern paper mills. In an industry like Kuantum Papers, where precise control over water, steam, and chemical usage is necessary to ensure product quality and operational efficiency, flow meters serve as the backbone of real-time monitoring systems. Every stage of paper manufacturing—from pulping and bleaching to drying and chemical dosing—requires accurate flow data to maintain consistent process conditions. The right flow measurement technology ensures that utilities are used optimally, losses are minimized, and production is uninterrupted.
At Kuantum Papers, different types of flow meters are deployed depending on the application, type of fluid, and required accuracy.

3.1.1 ORIFICE PLATE:

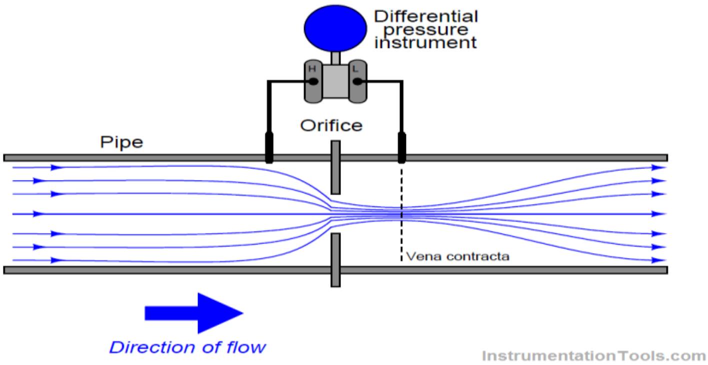
The Orifice Flow Meter is one of the most commonly used primary flow elements in industrial settings, and it plays a significant role in the process control systems at Kuantum Paper Mill. It is based on the Bernoulli's principle, where a fluid passing through a restriction (orifice) causes a measurable pressure drop. The orifice plate is a flat, circular plate with a sharp-edged hole in the center, installed between flanges in a pipeline. When fluid flows through this constriction, its velocity increases and pressure drops, creating a differential pressure that is directly proportional to the flow rate. This differential pressure is measured using a differential pressure transmitter connected to pressure taps upstream and downstream of the orifice. In paper mills, orifice meters are often employed in steam lines, cooling water circuits, and compressed air systems, where the fluid is relatively clean and free of suspended solids.

The basic working principle relies on **Bernoulli's theorem** and the **continuity equation**, with flow rate calculated using the formula:

$Q = Cd \times A \times \sqrt{(2\Delta P/\rho)}$

where Cd is the discharge coefficient, A is the area, ΔP is the pressure drop, and ρ is the fluid density. The **beta ratio ($\beta = d/D$)**, representing the ratio of orifice diameter to pipe diameter, typically ranges between 0.7 to 0.8 for optimal performance. In paper mills, orifice plates are widely used to measure **steam flow** due to their **low cost, simple design, and ruggedness**, although they are best suited for **clean fluids**. While orifice flow meters are less accurate than some advanced types, they remain popular due to their ease of use and reliability in steady-state operations.

For **orifice plates**, maintaining the sharp edge of the plate is vital. If the plate becomes **flattened or corroded**, the **vena contracta** (narrowest flow point) may not form correctly, resulting in incorrect differential pressure (ΔP) readings. It's recommended to inspect and **replace orifice plates every 10 years** or earlier if wear is observed.



3.1.2 ELECTROMAGNETIC FLOW METER (MAG):

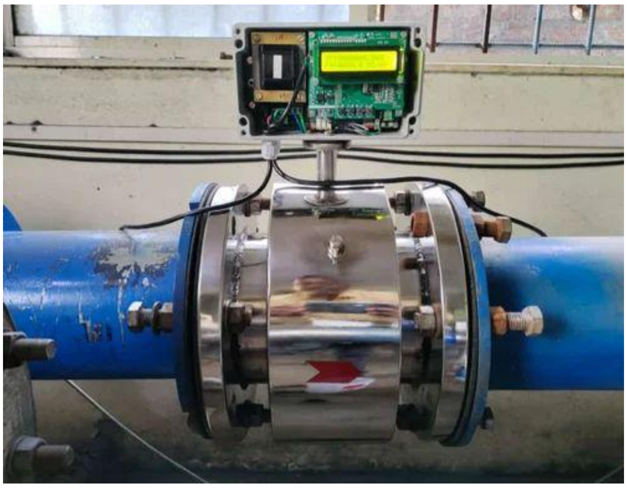
The electromagnetic (or Mag) flow meter is ideal for measuring the flow of **conductive fluids**, making it highly suitable for paper mill applications such as **pulp slurry, black liquor, white liquor, and chemical dosing**. It operates based on **Faraday's Law of Electromagnetic Induction**, which states that a voltage is induced when a conductive fluid moves through a magnetic field. This induced voltage is directly proportional to the flow velocity, described by the equation:

$$v = E / (B \times D),$$

where V is the fluid velocity, E is the induced voltage, B is the magnetic field strength, and D is the pipe diameter. The volumetric flow rate is calculated using

$$Q = v \times A = (\pi \times E \times D) / (4 \times B).$$

Mag meters offer **high accuracy (±0.5%)**, no moving parts, and are unaffected by temperature, pressure, or viscosity, making them well-suited for corrosive and slurry-type fluids. They are widely used in paper mills due to their **durability, low maintenance, and versatility** across various process stages, especially where fluid conductivity is sufficient. Moreover, electromagnetic flow meters provide **bidirectional flow measurement**, have **no pressure loss**, and can handle **large pipe diameters**, which is advantageous in large-scale industrial setups like digesters, bleaching plants, and stock preparation sections.



Proper maintenance of flow meters is crucial to ensure long-term accuracy and reliability in industrial applications. For **electromagnetic flow meters**, regular checking of **earthing connections** is essential to prevent **static currents** and interference in signal readings. The flow tubes and electrodes should be periodically cleaned to avoid **scaling or coating buildup**, especially during plant shutdowns. Materials passing through the sensors must be conductive; otherwise, the meter will not function correctly

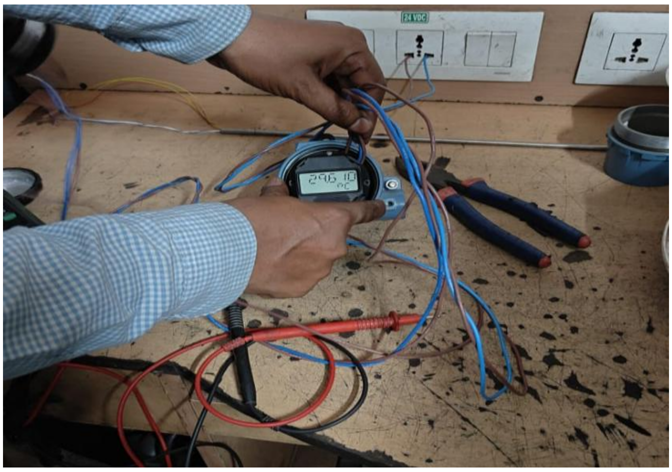
3.2 TEMPERATURE CONTROL AND MONITORING:

Temperature control and monitoring play a critical role in the efficient and safe operation of any process industry, especially in paper manufacturing where thermal processes are integral to product quality and energy optimization. In Kuantum Paper Mill, maintaining precise temperatures is essential in digestion, bleaching, drying, and chemical recovery sections, where variations can impact pulp strength, fiber bonding, drying efficiency, and chemical reaction rates. Accurate temperature monitoring ensures that each unit operation is carried out under ideal thermal conditions, reducing process deviations and minimizing energy waste. This is achieved through the use of RTDs (Resistance Temperature Detectors), thermocouples, transmitters, and control loops integrated into automated systems like DCS or PLC. With continuous monitoring, temperature data is used to make real-time adjustments, allowing for predictive maintenance, quality assurance, and optimized steam usage, all of which contribute to consistent product standards and cost-effective operations.

3.2.1 RESISTOR TEMPERATURE DETECTOR(RTD):

RTDs are widely used in industries for precise temperature measurement due to their high accuracy, stability, and repeatability. At Kuantum Paper Mill, RTDs—especially the Pt-100 type—are commonly used to monitor medium-range temperatures, typically up to **300°C.** The **Pt-100 RTD** is made from platinum and has a resistance of 100 ohms at 0°C, increasing in a known manner as temperature rises. This property allows for accurate and linear temperature readings, essential in maintaining process control in pulp and paper production. A popular configuration used is the **3-wire duplex RTD**, which minimizes errors caused by lead wire resistance and provides redundancy—two sensors in one sheath for simultaneous or backup readings. During the internship, students were shown how to wire RTDs correctly, identify terminals, and check for continuity using a multimeter. This hands-on experience emphasized the importance of proper connection techniques, as even small errors in wiring can lead to incorrect readings or process deviations. RTDs are typically connected to a temperature transmitter or directly interfaced with the DCS (Distributed Control System) to provide real-time monitoring and control.

A duplex RTD consists of **two sensing elements housed in the same probe**. This configuration is extremely useful in process industries, as it offers redundancy—if one element fails, the other continues to function, allowing uninterrupted monitoring. This is particularly valuable in high-dependency systems such as paper drying sections or steam control valves, where failure in temperature measurement could lead to compromised product quality or unsafe operating conditions. The second channel in a duplex RTD also allows for **simultaneous signal transmission to two different systems**, such as a PLC and an external display or safety interlock



3.2.2 THERMOCOUPLE:

At Kuantum Paper Mill, **K-type thermocouples** play a crucial role in measuring high-temperature areas, particularly in sections like steam headers, drying cylinders, and chemical processing units. These thermocouples are chosen for their robustness, broad **temperature range (-200°C to 1250°C)**, and fast response time, making them ideal for rapid thermal variations in process industries. The working principle is based on the **Seebeck effect**, where the temperature difference between two junctions of dissimilar metals (Chromel and Alumel) produces a small voltage signal, which is then converted into a readable temperature value. In field observations, an output of 1 millivolt corresponds to approximately 24°C, showing a reasonably linear behavior across typical operational ranges. A signal reading of 21.746 mA from the transmitter indicates a working temperature near the upper operating limits, and this current output is used in industry-standard 4–20 mA loops for integration with the DCS or PLC. The system employs duplex thermocouples, where two thermocouple elements are housed in one sheath. This redundancy ensures continuous monitoring—even if one channel fails, the other continues operation, thus enhancing safety and reliability. These sensors are calibrated routinely, and their wiring is checked to avoid polarity reversals or voltage drop errors, which could affect critical temperature-dependent operations in pulp drying or steam control. Accurate monitoring via thermocouples ensures that the temperature remains within operational limits, thus preventing process disturbances and optimizing product quality.

3.3 PRESSURE AND LEVEL MEASUREMENT TECHNIQUES:

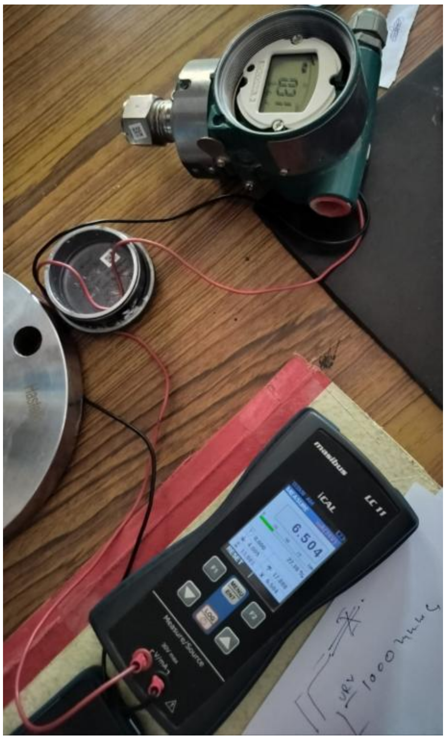
Pressure and level measurement are fundamental parameters in industrial automation, ensuring safe and efficient plant operation. In a paper mill, pressure measurement is crucial for controlling steam systems, chemical feed lines, and maintaining safe operating conditions in boilers and digesters. Similarly, level measurement ensures accurate monitoring of liquids in tanks, reservoirs, and reactors, preventing overflows, dry runs, and process interruptions. Devices such as pressure transmitters, differential pressure sensors, radar level sensors, and ultrasonic level transmitters convert physical quantities into electrical signals, allowing integration with PLC/DCS systems for real-time monitoring, control, and alarms. Together, these measurements play a critical role in maintaining process stability, equipment safety, and overall operational efficiency.

3.3.1 CAPACITIVE PRESSURE AND LEVEL TRANSMITTER:

In the paper industry, capacitive pressure and level transmitters play a crucial role in ensuring process stability and product quality. These transmitters operate on the principle that the capacitance between a diaphragm and an electrode changes with applied pressure or material level. A common application is in tanks and digesters where both pressure and level need continuous monitoring. The transmitter consists of a sensing diaphragm, which deflects in response to pressure or hydrostatic head (in case of level), and this deflection alters the capacitance value, which is then converted into a standard 4–20 mA current signal. For example, when the signal is 12 mA, it indicates that the process variable (pressure or level) is at 50% of its range. The range is defined between LRV (Lower Range Value) and URV (Upper Range Value); say, if LRV is 0 and URV is 100 cm, then 12 mA corresponds to a 50 cm level.

In pressure measurement, capacitive transmitters are installed in critical zones like pulp digesters, steam lines, and hydraulic press systems to monitor pressure variations accurately. These transmitters are often calibrated using HART communicators, allowing engineers to fine-tune the output range, zero/span, damping, and diagnostics. A diaphragm-based sensor element ensures accurate and stable readings even in harsh conditions like high humidity or corrosive pulp slurry. In some installations, the transmitter output reads, for example, 21.746 mA, indicating a value close to the maximum pressure range. This real-time data helps maintain safe operating conditions and prevents overpressure situations that can cause mechanical damage.

Similarly, for level measurement, the same capacitive principle is used in storage tanks or chemical dosing systems where continuous level monitoring is essential. The diaphragm senses the hydrostatic pressure created by the liquid column, and the transmitter converts it to a proportional current signal. If the tank is half-filled, the output will be around 12 mA. The system is also highly reliable due to its duplex configuration—meaning if one sensing channel fails, the other can continue to function. This redundancy enhances safety in critical operations. Photos of field-mounted transmitters often show the diaphragm assembly connected via capillaries to remote seals, especially in tall or pressurized tanks. Calibration with a HART communicator is again used to align the transmitter output with actual level conditions, ensuring high accuracy and repeatability in measurement.



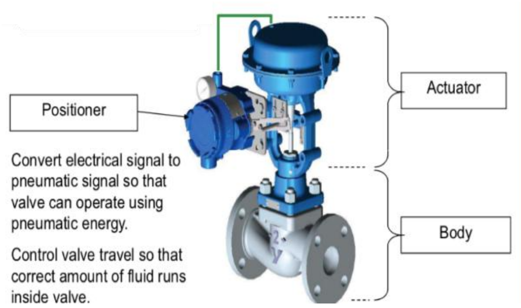
3.4 Field Caliberation Tools:

In modern process industries like paper manufacturing, **field calibration tools** are essential for maintaining the accuracy and reliability of instrumentation systems. One of the most commonly used tools is the **HART communicator (Highway Addressable Remote Transducer)**, which allows two-way digital communication with smart field instruments over the same 4–20 mA analog signal line. The HART communicator is widely used for the **calibration and configuration of RTDs, thermocouples, and pressure transmitters**. In RTD and thermocouple-based temperature transmitters, HART enables technicians to input temperature ranges, select sensor types (e.g., Pt100 or K-type), and fine-tune the output to match actual readings. For example, if the RTD reads 111 ohms corresponding to 26°C, or the thermocouple gives 1 mV at 24°C, the HART device ensures this data is correctly interpreted and transmitted as a linear analog signal.

In **pressure transmitters**, especially capacitive type used for level and pressure measurement, the HART communicator plays a crucial role in setting **Lower Range Value (LRV)** and **Upper Range Value (URV)**, checking output current (e.g., 12 mA for 50% level), and performing loop checks during commissioning. It also provides diagnostic information like sensor health, damping settings, and response time, making it invaluable for predictive maintenance. Whether adjusting a diaphragm-based pressure transmitter in a digester tank or verifying an RTD's output on a steam line, the HART communicator simplifies fieldwork, reduces downtime, and enhances system accuracy across various process points.

3.5 Difference between control valve and on/off valve:

In paper industries, especially around paper machines, we frequently observe both control valves and on/off valves integrated into the DCS (Distributed Control System). These valves serve distinct purposes in automation and flow management. **Control valves** are commonly used in critical process loops where continuous regulation of variables like steam flow, pressure, or chemical dosing is necessary. In DCS screens, they appear as modulating elements with real-time position feedback, often linked to analog signals such as 4–20 mA. Their role is to **regulate the flow dynamically**, adjusting their opening percentage based on PID controller outputs to maintain a setpoint. On the other hand, **on/off valves**, as seen in water line isolations or safety interlocks, operate in binary states—either fully open or fully closed. These are represented in DCS as simple actuators, triggered by digital signals, with no intermediate positioning. The **main difference** lies in their functionality: control valves provide **continuous, precise modulation** for process control, while on/off valves are used for **start-stop functions** without intermediate states. This differentiation is crucial in the paper manufacturing process, where fine control ensures product quality, and quick shutoff guarantees safety and efficiency.



3.6 Idea of Distributed Control System:

A **Distributed Control System (DCS)** is an advanced automated control platform used in industrial processes to monitor and control field instruments across large, complex operations. Unlike centralized control systems, DCS distributes control functions across multiple controllers, ensuring faster response, better reliability, and modular management. It integrates various sensors, transmitters, and actuators—displaying process data on operator screens in real time while enabling automated feedback control based on set parameters. DCS plays a critical role in industries like **paper manufacturing**, where precise control over temperature, pressure, level, flow, and chemical dosing is essential for both quality and safety.

During our internship at **Kuantum Paper Mill**, we were exposed to the DCS systems installed in the **paper machine section, pulp preparation area, and chemical recovery units**. We observed various **screens displaying real-time data** such as **setpoints (SP)** and **measured process values (PV)**. This helped us understand how the DCS continuously compares actual process values against the desired setpoints and generates control signals to adjust valves, pumps, or other final control elements to minimize deviations. This automatic correction ensures stable operation and product consistency. For instance, in the steam pressure control loop of the dryer section, the DCS maintained the required pressure by modulating the control valve as per the SP-PV difference.

One of the most insightful aspects of our learning was viewing the **trend analysis** provided by the DCS. These trends offered a graphical view of process parameters over time, allowing operators to detect anomalies, gradual drifts, or sudden spikes. We saw how operators used these trend lines to anticipate equipment failure or to fine-tune process efficiency.

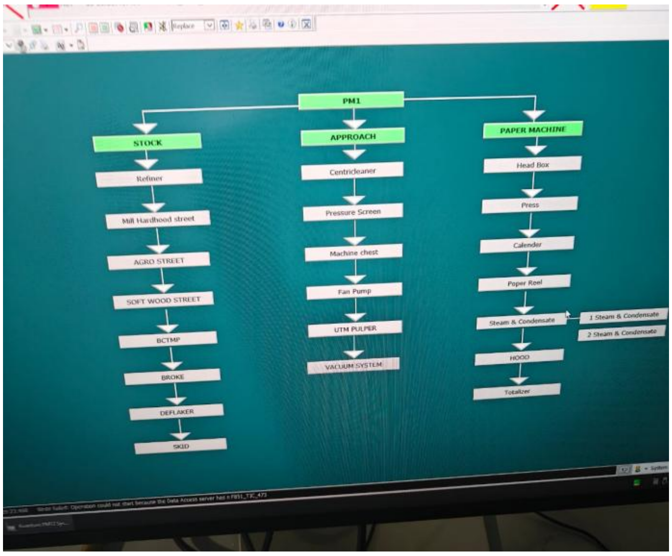


Figure 4 DCS Screen

Figure 6 DCS in Pulp Mill

Figure 5 DCS Screen

4. PROJECT: PREDICTIVE MODELLING OF SCALING PHENOMENON USING INDUSTRIAL SENSOR DATA

4.1 Problem Statement:

In the pulp and paper industry, scaling within process instrumentation—especially in bleaching stages—can lead to equipment inefficiencies, sensor inaccuracies, and increased maintenance downtime. These issues not only disrupt continuous operations but also elevate production costs and compromise product quality.

This project aims to develop a machine learning-based predictive model that analyses historical process data and identifies key parameters contributing to scaling, specifically in D0-stage sensors. By examining multiple process variables (X-factors) such as pH, flow rate, chemical concentrations, and temperature—recorded at regular intervals—the objective is to uncover patterns and conditions that typically precede scaling events. With limited data representing only the periods when scaling occurred, and corresponding cleaning dates logged manually, the challenge is to derive insights through data preprocessing, feature pattern analysis, and classification modelling.

4.2 What is Scaling?

Scaling is the undesirable deposition of solid materials—such as calcium carbonate, calcium oxalate, silica, or other inorganic salts—on the surfaces of pipelines, sensors, and equipment in the paper manufacturing process. Scaling in industrial systems, including paper mills, often begins with the formation of a small initial deposit known as a "nucleus." This nucleus acts as a seed or anchor point upon which further scale builds up rapidly. In the context of sensor surfaces or pipe walls, certain conditions—such as a sudden spike in pH, temperature, or concentration of dissolved minerals—can trigger the formation of this initial nucleus. Once the nucleus forms, it disrupts the local flow dynamics and promotes further precipitation of dissolved salts around it. As a result, the scaling process accelerates, transitioning from a slow, microscopic build-up to a fast, visible accumulation. This mechanism explains why certain instruments suddenly show abnormal behaviour or signal drift after operating stably for weeks. In paper mills, where chemical balance is delicate—especially during stages like D0 bleaching—the early detection of such nucleation-friendly conditions is critical. Predictive modelling and data analysis help identify these precursors, allowing for preventive action before the scaling reaches a level that demands cleaning or halts production.

Scaling in the D_0 stage of a paper mill is primarily caused by the precipitation of inorganic compounds such as calcium oxalate, calcium sulfate, and magnesium hydroxide, which tend to deposit on equipment surfaces under certain chemical and thermal conditions. One key contributor is an abrupt shift in pH — especially a transition from acidic to slightly basic — which can decrease the solubility of metal ions and trigger rapid crystallization. Additionally, overdosing of chlorine dioxide (ClO_2), elevated temperatures, and poor control of brightness chemicals can exacerbate scaling. These conditions encourage nucleation and crystal growth, especially if the mixing is uneven or if residual metal ions from previous stages are present. Careful regulation of pH, ClO_2 dosage, and temperature is thus essential to prevent scaling and maintain efficient bleaching operations.

Figure 7 Scaling in Brightness Sensor

Figure 8 Scaling in Brightness Sensor

4.3 Theoretical effect of all variables:

No.	Variable	Description	Scaling Risk Summary
1	ClO_2 Treated Pulp	Bleached pulp quality	Poor bleaching → ↑ scaling
2	H_2SO_4 Flow	Acid for pH control	Fluctuation → corrosion/scale
3	Feed Line Kappa	Lignin content in pulp	High Kappa → ↑ scaling
4	Brightness (MC Pump)	Pulp whiteness after O_2 stage	Low brightness → ↑ residuals
5	Effluent Temp (Cooler)	Hot water temp in acid cooler	High temp → ↑ chemical scaling
6	TRP Production Rate	Pulp flow rate	Very high/low flow → ↑ scaling
7	Washer pH	pH post washing	Low pH → metal redeposition
8	O_2 Reactor Temp	Temp after oxygen reactor	High temp → precipitate formation
9	O_2 to ODL Valve	Oxygen control to delignification	Instability → fouling
10	NaOH to Repulper	Alkali for delignification	Imbalance → soap or poor cleaning
11	ClO_2 Flow	Chlorine dioxide dosage	Unbalanced dose → scale/corrosion
12	Inlet pH	pH before D_0 stage	Extreme pH → sensor/chemical scaling
13	Outlet pH	pH after D_0 stage	Drop indicates possible in-line scaling
14	Washer NaOH Flow	Alkali in washing	High/low flow → incomplete neutralizing
15	SO_2 Flow	Acidification agent	High flow → sulphate/corrosion scaling

4.4 Objectives:

- To identify critical variables that contribute to scaling in D0-stage sensors.
- To clean and preprocess historical process data.
- Analysis using Orange Data Mining
- To develop a logistic regression model to classify scaling events.
- To evaluate model performance and accuracy.
- To generate insights for proactive scaling prevention.

4.5 Tools and Libraries:

- **Logistic Regression:**

Logistic regression is a simple yet powerful tool used to predict the chances of something happening — especially when the outcome is either a "yes" or a "no." In our case, we're using it to figure out whether or not scaling is likely to occur in a paper mill's D0 pH sensor based on certain measurable factors like pH, temperature, flow rate, and more.
Unlike regular linear regression, which predicts numbers like temperature or pressure values, logistic regression predicts the probability of an event — like scaling. It uses a special function called the sigmoid function, which takes all the input factors and gives us a probability between 0 and 1. If the probability is above a certain threshold (usually 0.5), we consider that scaling is likely to happen.
The core formula looks like this:
$P(\text{scaling}) = 1 / (1 + e^{\wedge}- (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta n X n))$
Here, the X values represent our input features (like pH or flow), and the β values are weights that the model learns based on past data. These weights tell us how important each feature is in contributing to the final prediction.
To measure how well our logistic regression model fits the data, we use something similar to R^2 from linear regression, but it's called McFadden's R^2. While it doesn't behave exactly the same, it helps us understand how well the model explains the outcome. A McFadden's R^2 value between 0.2 and 0.4 is considered quite good. In our analysis, the model gave us a McFadden's R^2 of approximately [insert value], which means it does a reasonably good job at explaining when scaling might occur.
What makes logistic regression especially useful in our context is that it doesn't just make predictions — it also helps us understand which factors matter the most. For example, if the model gives a high weight to the inlet pH, it tells us that changes in pH strongly influence the chances of scaling. This kind of insight can help engineers take proactive steps to avoid scaling and improve the performance of equipment.
Overall, logistic regression gives us both a predictive model and a deeper understanding of what's going on in the system — making it a valuable part of our data-driven approach in the paper mill.

- **Orange Data Mining:**

Orange is an open-source, Python-based data analysis and machine learning platform that allows users to build visual workflows using a simple drag-and-drop interface. It is particularly helpful for quickly exploring data, building models, and interpreting results without writing code.

In our project, Orange was used to analyse sensor data from the paper mill to identify the factors leading to scaling in the D0 stage. We utilized several key widgets, including Data Table for data inspection, Box Plot for visualizing feature distributions, Rank for identifying the most influential variables, and Logistic Regression for building a predictive model. These tools enabled us to explore the dataset, observe trends, and evaluate feature importance effectively.

The Logistic Regression widget allowed us to predict the likelihood of scaling events based on selected process variables. Model performance was evaluated using built-in tools like the confusion matrix and classification accuracy. Orange's interactive and visual approach helped us understand complex data patterns, making it easier to communicate insights and draw meaningful conclusions for early detection and prevention of scaling.

- **Python Libraries:**

pandas, sklearn.linear_model, sklearn.model_selection, sklearn.preprocessing, sklearn.metrics

4.5 Steps Involved during the project:

- **Data Collection**

The data for this project was collected from a paper mill, focusing on the D0 stage of the bleaching process where scaling often affects instrument performance. We gathered two types of data:

a) Sensor/Process Variable Data (X-values):
Approximately 4 months of time-series data recorded at 15-minute intervals was collected from various sensors in the D0 stage. These included parameters like pH (inlet and outlet), flow rate, conductivity, temperature, kappa number, brightness, and pressure. This data provided insights into the process conditions leading up to scaling events.

b) Scaling Event Data (Y-values):
Scaling-related events were not digitally recorded, so we manually extracted cleaning dates from operator logbooks. These logs indicated when instruments, especially the D0 pH sensor, were cleaned due to suspected scaling. We used these dates to label portions of the dataset that likely represented scaling conditions.

- **Data Cleaning**

Once the raw data was collected, the next critical step was data cleaning to ensure accuracy, consistency, and suitability for machine learning analysis. Given that the sensor data was recorded at 15-minute intervals over several months, it required careful preprocessing.

a) Handling Duplicates and Redundancy:
We checked for duplicate rows and overlapping entries across multiple sheets. In some cases, multiple entries had the same timestamps due to system errors or data logging issues, which were resolved by averaging or keeping only the first occurrence per time window.

b) Outlier and Missing Value Treatment:
Although there were minimal missing values, we reviewed the dataset for anomalies—such as sudden spikes or drops in pH or temperature—which could indicate sensor faults rather than genuine process behaviour. Where necessary, such outliers were either imputed or excluded.

c) Scaling Label Creation:
Since direct labels of scaling were unavailable, we used the manually extracted cleaning dates to create a binary target column. For each recorded cleaning date, we assumed that scaling had

occurred during the previous 1 week, and marked those rows as '1' (scaling). The rest were marked as '0' (no scaling). This target variable was then used in supervised learning models.

d) Consolidation of Sheets:
Sensor variables were distributed across different sheets. We combined the relevant columns into a single master sheet, aligned by timestamp, to create a complete dataset ready for further analysis.

Overall, data cleaning transformed unstructured and partially labelled data into a reliable and consistent dataset. This step ensured the integrity of the model training and helped improve the accuracy of pattern detection and predictions.

- **Manual Analysis**

Before applying any machine learning algorithms, we performed manual analysis to understand the behaviour of individual variables and their possible role in scaling. This step helped in identifying patterns and building domain intuition about the dataset.

a) Trend Observation:
We visually inspected the time-series data of each process variable (such as inlet pH, outlet pH, flow rate, kappa number, temperature, etc.) around the dates when scaling was known to have occurred. This helped us hypothesize which parameters showed abnormal changes or drifts leading up to scaling.

b) Statistical Summary:
Basic statistics like mean, median, and standard deviation were computed for each feature, particularly comparing values during scaling and non-scaling periods. This gave a preliminary sense of which features might be more discriminative.

c) Correlation Checks:
We also explored the correlation between different variables to see if any particular pair showed a strong relationship—either positively or negatively. For example, a drop in outlet pH combined with a spike in conductivity could indicate precursor conditions for scaling.

d) Domain Expert Input:
With the guidance of mill staff and mentors, we learned about the chemical mechanisms behind scaling and the operational significance of certain parameters. This helped validate and refine our assumptions before formal modelling.

This manual step was important to guide the machine learning workflow, especially in selecting relevant features, understanding data behaviour, and verifying whether observed patterns made sense from a chemical and operational perspective.

- **Orange Model**

After manual analysis, we used Orange Data Mining to visually model and analyse our dataset. Orange's drag-and-drop interface enabled rapid experimentation and better interpretability of our features and results.

a) Data Preparation and Input:
We imported the cleaned dataset, ensuring all sensor variables were set as input features and the scaling column was marked as the target.

b) Feature Ranking and Visualization:
Using the Rank widget, we identified the most influential variables, such as inlet pH, outlet pH,

and kappa number. Box plots and scatter plots helped us observe variations and relationships between features under scaling and non-scaling conditions.

c) Model Development and Evaluation:
We applied logistic regression using Orange's widget and evaluated it using the Test & Score and Confusion Matrix widgets. The model's accuracy and AUC helped validate its predictive strength.

d) Data Table Insight:
The Data Table widget allowed us to inspect individual records to understand prediction correctness and refine our approach.

This phase helped us build an interpretable base model and explore key patterns in the data before moving to Python-based modelling.

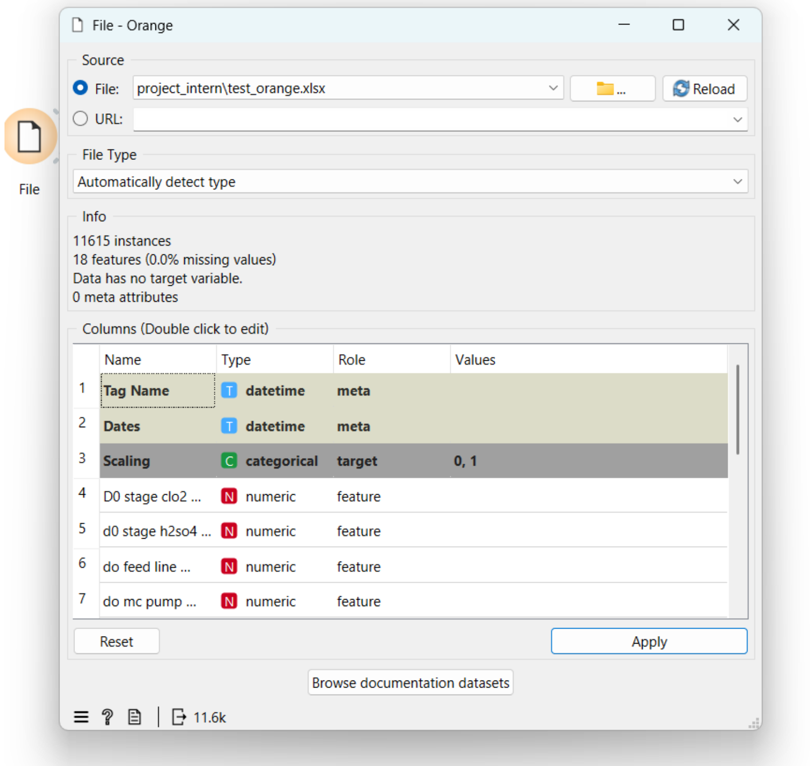


Figure 9 Uploading file in Orange

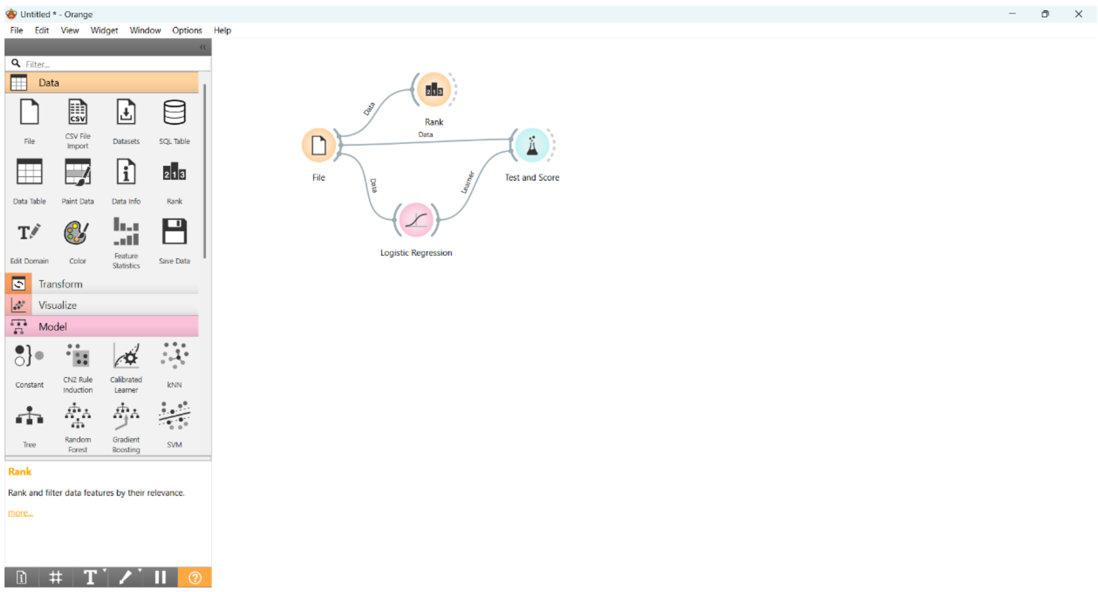


Figure 10 Orange Drag and Drop Interface

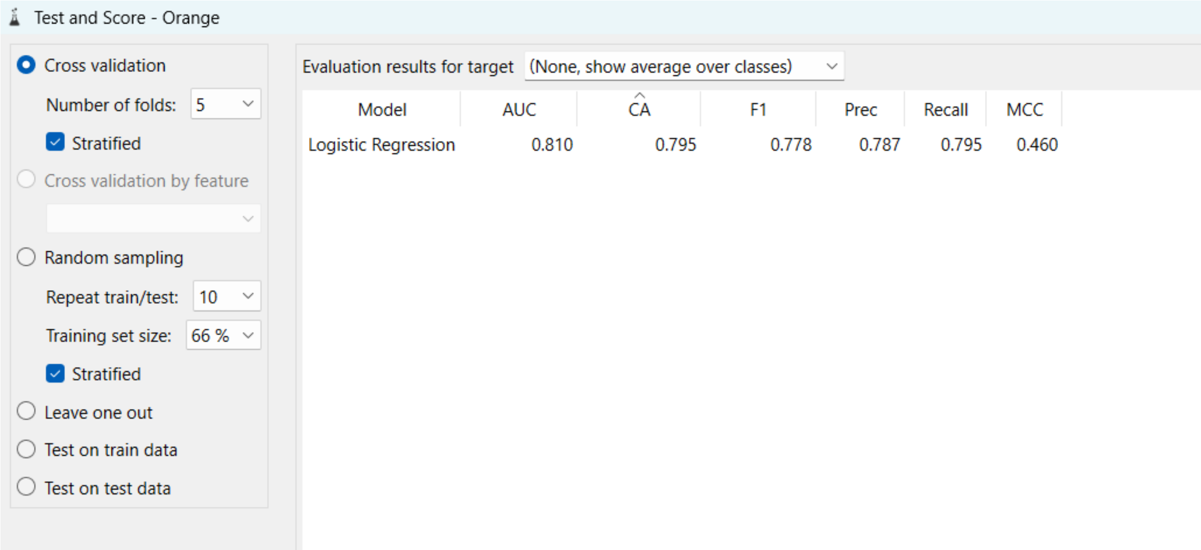


Figure 11 Orange Logistic Regression Model

Machine Learning Model Using Python

Once the data had been cleaned and manually analysed, we transitioned to building a machine learning model using Python to predict scaling in the D_0 sensor of the paper mill. The goal was to develop a predictive framework capable of identifying patterns that precede scale formation, using historical sensor data.

The logistic regression algorithm from the scikit-learn (sklearn) library was selected due to its suitability for binary classification problems and its interpretability. We began by importing the cleaned dataset into Python using the pandas library. The 'Scaling' column, manually created based on expert-verified cleaning dates, served as the target variable. The remaining 15 sensor-based variables (such as inlet pH, outlet pH, kappa number, etc.) were treated as the feature set. Before model training, non-numerical and irrelevant columns (like timestamp or remarks) were removed, and all feature variables were normalized using the StandardScaler. This step ensured that each variable contributed equally to the model and removed any bias due to varying units or scales.

The dataset was then split into training and testing subsets using train_test_split, with 33% of the data reserved for model evaluation. The logistic regression model was trained on the training set using sklearn's Logistic Regression function. After training, the model coefficients and intercept were extracted to construct the logistic regression equation, which illustrates how each sensor variable affects the likelihood of scaling. For example, a positive coefficient for inlet pH would indicate that higher pH levels are associated with an increased risk of scaling.

To assess the performance of the model, we calculated the accuracy score, which reflects the proportion of correctly predicted scaling and non-scaling instances in the test set. Additionally, we computed McFadden's pseudo-R^2 — a metric used to assess the goodness of fit for logistic regression models — which gave us a better sense of how well our model explained the variability in the target variable.

The final model not only provided satisfactory accuracy but also offered interpretability through its coefficients. This allowed us to rank the contributing variables based on their influence on scaling, enabling actionable insights. Through this Python-based machine learning approach, we established a robust, data-driven method for early detection of scaling conditions, offering a valuable tool for preventive maintenance and operational optimization in the pulp mill.

Result:

The logistic regression model achieved an accuracy of 80.30% with a McFadden's pseudo-R^2 of 0.1960, indicating a reasonably good fit for predicting scaling in the D_0 stage. Key variables such as D_0 stage ClO_2-treated pulp, acid effluent cooler temperature, and washer pH showed strong positive influence on scaling, while factors like H_2SO_4 flow, outlet pH, and MC pump brightness had a negative impact, reducing scaling risk. These insights help identify critical process parameters for controlling and minimizing scale formation in the pulp mill.

Python Code:

```python
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, log_loss
import numpy as np

# Step 1: Load the Excel file
df = pd.read_excel("test_orange.xlsx")

# Step 2: Split target (y)
y = df["Scaling"]

# Step 3: Drop target and select only numeric features
X = df.drop(columns=["Scaling"])
X_numeric = X.select_dtypes(include='number')  # Remove date/time and text columns

# Step 4: Scale features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_numeric)

# Step 5: Train/test split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.33, random_state=42)

# Step 6: Train Logistic Regression
model = LogisticRegression()
model.fit(X_train, y_train)

# Step 7: Print model equation
coefficients = model.coef_[0]
intercept = model.intercept_[0]
feature_names = X_numeric.columns

print("\nLogistic Regression Equation:")
equation = "logit(p) = {:.3f}".format(intercept)
for coef, fname in zip(coefficients, feature_names):
    equation += " + ({:.3f} * {})".format(coef, fname)
print(equation)
print("\nProbability(p) = 1 / (1 + exp(-logit(p)))")

# Step 8: Evaluate accuracy
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("\nModel Accuracy: {:.2f}%".format(accuracy * 100))

# Step 9: McFadden's pseudo R²
ll_full = -log_loss(y_test, model.predict_proba(X_test), normalize=False)
ll_null = -log_loss(y_test, [np.mean(y_test)] * len(y_test), normalize=False)
mcfadden_r2 = 1 - (ll_full / ll_null)
print("McFadden's pseudo R²: {:.4f}".format(mcfadden_r2))
```

Output:

```
Logistic Regression Equation:
logit(p) = -1.297 + (1.541 * D0 stage clo2 treated pulp) + (-0.558 * d0 stage h2so4 flow) + (-0.052 * do feed line kappa) + (-0.647 * do mc pump outlet brightness) + (0.675 * Acid effluent cooler hot water temperature control process value) + (-0.040 * trp feed production rate) + (0.866 * hw d0 washer ph) + (-0.384 * o2 reacter delivery tt) + (-0.169 * o2 to odl fcv process value) + (0.009 * caustic flow to odl repulper) + (0.066 * clo2 flow) + (-0.376 * d0 inlet ph) + (-1.323 * d0 outlet ph) + (-0.043 * d0 washer naoh flow) + (0.377 * do so2 flow)

Probability(p) = 1 / (1 + exp(-logit(p)))

Model Accuracy: 80.30%
McFadden's pseudo R²: 0.1960
```

4.6 Recommendations

Based on the machine learning model's insights, several key process parameters were identified as significant contributors to scaling in the D_0 stage, and controlling these factors can effectively reduce its occurrence. Brightness plays a vital role—lower brightness levels often point to inefficient bleaching and chemical imbalances that can lead to scale deposits. Chlorine dioxide (ClO_2) flow must also be carefully managed; inconsistent dosing can disrupt the oxidation reactions, allowing scaling compounds to form. Temperature, if too high or fluctuating, accelerates the rate of precipitation and crystallization reactions, thus promoting scale buildup. Moreover, pH emerged as a crucial variable—an excessively low or high pH can shift the solubility of dissolved solids, leading to their deposition on sensor and pipeline surfaces. To mitigate these issues, it is recommended that the pulp mill maintain optimal and stable ranges of brightness, ClO_2 flow, temperature, and pH. Incorporating automated control systems and real-time monitoring of these variables can enable early detection of scaling-prone conditions, thereby improve overall process efficiency and reduce downtime due to maintenance. By focusing on these critical variables, the pulp mill can minimize the frequency and severity of scaling incidents, enhance operational stability, and reduce unplanned maintenance.

5. CONCLUSION:

This internship has been a culmination of multiple layers of learning — combining process engineering, instrumentation, chemical theory, industrial practices, and data analytics — applied to understand and predict scaling in the D_0 stage of a paper mill. Through the course of this work, we gained an in-depth understanding of the pulp bleaching process, particularly the functioning of the D_0 stage, where chlorine dioxide (ClO_2) is used under acidic conditions to remove residual lignin from pulp. We explored how variations in pH, temperature, chemical dosages, and the presence of dissolved ions contribute to the formation of scale — a common and problematic occurrence in industrial pipelines and sensors.

On the instrumentation front, we learned how field sensors such as pH probes, flow meters, and brightness analyzers operate in harsh environments and how their readings are affected by scale deposition. The challenges of sensor fouling, signal drift, and inaccurate readings were studied in detail, helping us appreciate the role of preventive maintenance and real-time monitoring. The importance of accurate and consistent sensor data was emphasized throughout the project, especially while labeling data and training machine learning models.

Our mentor also introduced us to industrial process systems like black liquor recovery, chemical dosing loops, heat exchangers, and washing systems — expanding our understanding of unit operations in a working pulp mill. From observing the flow of black liquor through evaporators and recovery boilers to tracing white liquor regeneration in the causticizing plant, we connected theory with practical insights. These experiences highlighted the interdependence of process efficiency, instrumentation accuracy, and chemical balance in maintaining stable operations.

From a data science perspective, we developed strong foundational skills in cleaning time-series industrial data, selecting meaningful features, and applying machine learning algorithms for prediction. Orange Data Mining enabled us to visualize and explore patterns using tools like Rank, Box Plot, and Logistic Regression in an intuitive way. Python gave us the flexibility to build a logistic regression model with measurable accuracy and interpretability. We learned how to create a labeled dataset using cleaning dates, normalize variables, evaluate model performance, and generate equations that link process conditions to scaling risk.

In conclusion, this training went far beyond building a machine learning model. It enriched our understanding of how chemical processes behave, how instrumentation responds in real conditions, how data must be handled for accuracy, and how analytics can guide industrial decision-making. This integrated learning experience has prepared us to think like engineers — not just in isolated domains, but across disciplines — enabling us to contribute meaningfully to process industries, smart manufacturing, and sustainable engineering solutions in the future.