**Industrial Internship Report on**

“Multi-Stage Continuous-Flow Manufacturing Process – Predicting Factory Output”

**Prepared by:**

Riddhi Jangid

|  |
| --- |
| *Executive Summary* |
| This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).  This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time.  My project was (Tell about ur Project)  This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship. |

**TABLE OF CONTENTS**

[1 Preface 3](#_Toc139702806)

[2 Introduction 4](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 4](#_Toc139702808)

[2.2 About upskill Campus 8](#_Toc139702809)

[2.3 Objective 9](#_Toc139702810)

[2.4 Reference 9](#_Toc139702811)

[2.5 Glossary 10](#_Toc139702812)

[3 Problem Statement 11](#_Toc139702813)

[4 Existing and Proposed solution 12](#_Toc139702814)

[5 Proposed Design/ Model 13](#_Toc139702815)

[5.1 High Level Diagram (if applicable) 13](#_Toc139702816)

[5.2 Low Level Diagram (if applicable) 13](#_Toc139702817)

[5.3 Interfaces (if applicable) 13](#_Toc139702818)

[6 Performance Test 14](#_Toc139702819)

[6.1 Test Plan/ Test Cases 14](#_Toc139702820)

[6.2 Test Procedure 14](#_Toc139702821)

[6.3 Performance Outcome 14](#_Toc139702822)

[7 My learnings 15](#_Toc139702823)

[8 Future work scope 16](#_Toc139702824)

9 Appendix…………………………………………………………………………………………………………………………………………17

# 

# Preface

The Summer Internship at **Upskill Campus** provided me with an excellent opportunity to gain practical exposure in the field of **Data Science and Machine Learning**. During this internship, I worked on an industry-oriented project titled “Multi-Stage Continuous-Flow Manufacturing Process – Predicting Factory Output”.

This project focused on handling a real-world manufacturing dataset and applying predictive modeling techniques to forecast factory outputs. Through this experience, I learned how to preprocess raw industrial data, manage noisy measurements, and design machine learning models for multi-output prediction.

The internship not only enhanced my **technical proficiency in Python, Pandas, and Scikit-learn** but also gave me valuable insights into how machine learning can be integrated into industrial processes for performance monitoring and anomaly detection.

Working on this project has been an enriching experience, bridging the gap between academic knowledge and industrial application. I believe that the skills and learnings acquired during this internship will significantly contribute to my future endeavors in the field of Artificial Intelligence and Data Science.

# Introduction

## About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various**Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end**etc.



1. UCT IoT Platform **(****)**

**UCT Insight** is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

* It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
* It supports both cloud and on-premises deployments.

It has features to  
• Build Your own dashboard  
• Analytics and Reporting  
• Alert and Notification  
• Integration with third party application(Power BI, SAP, ERP)  
• Rule Engine



1. **Smart Factory Platform (****)**

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

* with a scalable solution for their Production and asset monitoring
* OEE and predictive maintenance solution scaling up to digital twin for your assets.
* to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
* A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

 

1.  based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

1. Predictive Maintenance

UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



## About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

<https://www.upskillcampus.com/>

upSkill Campus aiming to upskill 1 million learners in next 5 year



## 

## Objectives of this Internship program

The objective for this internship program was to

 ☛ get practical experience of working in the industry.

 ☛ to solve real world problems.

 ☛ to have improved job prospects.

 ☛ to have Improved understanding of our field and its applications.

 ☛ to have Personal growth like better communication and problem solving.

**Reference**

[1] Scikit-learn Documentation – *https://scikit-learn.org/*

[2] Pandas Documentation – *https://pandas.pydata.org/*

[3] Upskill Campus Learning Resources – [*https://upskillcampus.com/*](https://upskillcampus.com/)

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
| **Preprocessing** | The process of cleaning, transforming, and normalizing raw data before applying machine learning models. |
| **Setpoint** | The target value in a manufacturing process against which actual measurements are compared. |
| **Features (Inputs)** | Independent variables such as environmental conditions, machine parameters, and raw material properties used to train the predictive model. |
| **Labels (Outputs)** | Dependent variables or actual measured values that the model aims to predict. |
| **Random Forest Regressor** | A machine learning algorithm that uses an ensemble of decision trees to make predictions with improved accuracy. |
| **MultiOutputRegressor** | A technique that allows handling multiple output variables simultaneously in regression tasks. |
| **Mean Squared Error (MSE)** | An evaluation metric that measures the average squared difference between predicted and actual values. |
| **Anomaly Detection** | Identifying unusual patterns in manufacturing outputs by comparing predictions with actual values. |
| **StandardScaler** | A data normalization technique from Scikit-learn used to standardize features by removing the mean and scaling to unit variance. |
| **Continuous Flow Process** | A type of manufacturing system where materials are continuously fed through multiple stages without interruption. |

# Problem Statement

In modern manufacturing industries, **continuous-flow processes** are widely used for high-speed, large-scale production. These processes involve multiple machines operating in both **parallel and series stages**, making the system highly complex and interdependent.

The dataset provided for this project originates from a real factory production run, where:

* In the **first stage**, Machines 1, 2, and 3 operate in parallel and their outputs are combined.
* The **combined output** is measured at 15 different locations, referred to as **primary measurements**.
* In the **second stage**, Machines 4 and 5 operate in series, and the outputs are again measured at the same 15 locations, referred to as **secondary measurements**.

The key challenges identified are:

1. **High Dimensionality & Noise** – Measurements contain noise, making it difficult to accurately predict outputs.yes
2. **Multiple Outputs** – The model needs to predict **15 different output variables** simultaneously.
3. **Real-time Relevance** – Predictions should be efficient enough to be used in **real-time monitoring and anomaly detection**.
4. **Setpoint Tracking** – Each measurement has a setpoint, and the goal is not only to predict actual values but also to analyze deviations from setpoints.

Thus, the problem addressed in this project is:

**“To develop a predictive modeling framework that can utilize real process data from a multi-stage continuous-flow manufacturing line to accurately forecast the factory outputs and assist in real-time process monitoring and anomaly detection.”**

# Existing and Proposed solution

**4.1 Existing Solutions**

* Traditional rule-based monitoring systems are often inaccurate in noisy environments.
* Manual inspection delays decision-making.
* Generic ML models fail to capture complex process dependencies.

**4.2 Proposed Solution**

This project proposes a **machine learning–based predictive model** that:

* Build a **machine learning regression model** to predict 15 outputs.
* Use **Random Forest with MultiOutputRegressor** for better handling of multi-target regression.
* Apply **feature scaling and preprocessing** to clean noisy data.

## 4.3 Value Addition:

## Accurate predictions help optimize real-time process control.

## Can be extended to anomaly detection and predictive maintenance.

### ****4.4 Code Submission (GitHub link)****

[GitHub Repository - upskillCampus](https://github.com/riddhi844/upskillCampus)

### ****4.5 Report Submission (GitHub link)****

[GitHub Repository - upskillCampus](https://github.com/riddhi844/upskillCampus)

# Proposed Design/ Model

The proposed system is designed to build a **predictive framework** for a continuous-flow manufacturing process. It leverages **data preprocessing** and **machine learning models** to predict primary and secondary outputs.

**5.1 High Level Diagram**

**Description:**

* **Input Stage:** Environmental variables, raw material properties, and process variables from Machines 1–3.
* **Combiner:** Merges outputs of Machines 1–3.
* **Primary Output:** Measurements from 15 sensors (to be predicted).
* **Second Stage:** Machines 4–5 in series, further processing combined output.
* **Secondary Output:** Another set of 15 sensor measurements.
* **Machine Learning Model:** Random Forest with MultiOutputRegressor predicting 15 outputs simultaneously.

**5.2 Low Level Diagram**

**Step-wise flow:**

1. **Data Collection** – Load manufacturing\_data.csv (1 Hz sampling).
2. **Preprocessing** – Clean noisy data, drop constants, standardize features.
3. **Feature Extraction** – Inputs: Environment + Machines 1–3 + Combiner.
4. **Model Training** – Random Forest with MultiOutputRegressor.
5. **Evaluation** – Mean Squared Error (MSE) calculated for 15 outputs.
6. **Deployment** – Save model & scaler for real-time predictions.

**5.3 Interfaces**

The project currently uses **Python scripts** for implementation. Interfaces include:

* **Preprocessing Module:** Reads CSV, cleans data, scales features.
* **Training Module:** Builds and evaluates the ML model.
* **Prediction Module:** Loads model and scaler for new predictions.
* **Results:** Mean Squared Error (MSE) for each output, stored in results directory.

# Performance Test

**6.1 Test Plan**

The objective was to evaluate the accuracy of the model in predicting 15 primary outputs. Tests included:

* Dataset loading and preprocessing check.
* Feature scaling verification.
* Model training and prediction on unseen data.
* Error measurement using Mean Squared Error (MSE).

**6.2 Test Procedure**

* Preprocessed data using StandardScaler.
* Trained a **Random Forest + MultiOutputRegressor** model.
* Split data: **80% training, 20% testing**.
* Evaluated predictions against actual values using **MSE**.

**6.3 Performance Outcome**

* Model trained successfully on ~14,000 samples.
* **MSE values** per output ranged between **2.4 – 3.6**.
* More stable outputs had **lower errors**, while noisy ones showed **slightly higher MSE**.
* The model is suitable for **real-time monitoring** and **anomaly detection**.

# My learnings

During this internship project, I gained valuable exposure to both **technical and practical aspects** of data-driven manufacturing systems. My key learnings include:

* **Data Handling & Preprocessing**: Learned how to clean, normalize, and structure real-world noisy industrial data for machine learning.
* **Machine Learning Models**: Understood how to implement and train models like **Random Forest with MultiOutputRegressor** for multi-output regression tasks.
* **Model Evaluation**: Gained experience in using metrics like **Mean Squared Error (MSE)** to assess predictive performance.
* **Modular Coding**: Practiced building a structured project with preprocessing, training, and prediction modules.
* **Industry Relevance**: Understood how predictive models can be applied to **real-time process monitoring** and **anomaly detection** in manufacturing.

This project not only improved my **Python and ML skills** but also gave me confidence in applying theory to a **practical industrial scenario**.

# Future work scope

Although the current project successfully developed a predictive model for factory outputs, there is significant scope for future improvements and extensions:

* **Secondary Output Prediction**: Extend the model to handle the more **noisy measurements** from the second stage (Machines 4 and 5).
* **Real-Time Integration**: Deploy the trained model into a **real-time production environment** for live monitoring and anomaly detection.
* **Advanced Models**: Experiment with deep learning approaches (e.g., **LSTM or GRU networks**) to better capture time-series dependencies in the manufacturing process.
* **Anomaly Detection System**: Build an automated alerting system to flag deviations from setpoints beyond acceptable thresholds.
* **Visualization Dashboard**: Create an interactive dashboard for factory operators to view predictions, errors, and process trends in real time.
* **Scalability**: Optimize the system for **large-scale industrial deployment** with higher-frequency data.

**9 Appendix: Weekly Progress Reports**

**9.1 Week 1- Progress Report**

**Name:** Riddhi Jangid  
**Domain:** Machine Learning / Manufacturing Process Prediction  
**Week Ending:** 17 May 2025

**Overview:**  
Focused on understanding the manufacturing process dataset and setting up the project environment.

**Achievements:**

* Understood dataset structure (Machines 1–5, Combiner, primary & secondary outputs).
* Preprocessed data: cleaned, normalized, and prepared features/labels.
* Implemented preprocessing.py to handle raw input data.

**Challenges:**

* Dataset was noisy and had mixed data types.
* Faced issues while importing libraries and managing file paths.

**Learning Resources:**

* Pandas and Scikit-learn documentation.
* Internship-provided material.

**Next Week’s Goals:**

* Train predictive models for primary outputs.
* Evaluate performance using error metrics.

## 9.2 Week 2 - Progress Report

**Week Ending:** 24 May 2025

**Overview:**  
Implemented model training and evaluation for primary outputs.

**Achievements:**

* Trained a **Random Forest + MultiOutputRegressor** model.
* Achieved stable Mean Squared Error (MSE) results on 15 outputs.
* Saved trained model and scaler in results folder.

**Challenges:**

* Directory errors (missing results folder) caused saving issues.
* Managed large CSV size with proper preprocessing.

**Learning Resources:**

* Sklearn regression tutorials.
* GitHub examples of multi-output regression.

**Next Week’s Goals:**

* Build prediction script to test new inputs.
* Document results and performance evaluation.

**9.3 Week 3 Progress Report**

**Week Ending:** 31 May 2025

**Overview:**  
Covered basics of **Probability and Statistics** to strengthen theoretical foundation for ML.

**Achievements:**

* Studied probability distributions, mean, variance, and correlation.
* Related statistical measures with dataset patterns.
* Gained deeper understanding of model evaluation metrics (MSE, variance).

**Challenges:**

* Some statistical terms were difficult to link with practical dataset initially.

**Learning Resources:**

* NPTEL lectures on Probability & Statistics.
* Reference book: *Probability and Statistics for Engineers*.

**Next Week’s Goals:**

* Finalize performance testing and evaluation.
* Start drafting the internship report.

**9.4 Week 4 - Progress Report**

**Week Ending:** 7 June 2025

**Overview:**  
Focused on testing, final evaluation, and preparing report.

**Achievements:**

* Conducted performance tests on trained model with ~14,000 samples.
* Documented test plan, procedure, and outcomes (MSE range: 2.4 – 3.6).
* Completed structured internship report with Preface, Introduction, Methodology, Results, and Future Work.

**Challenges:**

* Balancing theoretical report writing with coding tasks.
* Ensuring report matches internship-provided format.

**Learning Resources:**

* Internship sample report.
* Online resources on ML in manufacturing.

**Next Week’s Goals:**

* Submit final GitHub repository with code and report.
* Share final report link on Google Form for evaluation.