AVOIDING SHORT SHOT QUALITY FAILURE IN INJECTION MOLDING MACHINE

A Thesis

by

Oğuz Yasin

Submitted to the

Graduate School of Sciences and Engineering
In Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in the
Department of Industrial Engineering

Özyeğin University
June 2021

AVOIDING SHORT SHOT QUALITY FAILURE IN INJECTION MOLDING MACHINE

Approved by:

Asst. Prof. Erinç Albey, Advisor,
Department of Industrial Engineering
Özyeğin University

Assoc. Prof. O. Örsan ÖZENER,
Department of Industrial Engineering
Özyeğin University

Assoc. Prof. M. Güray Güler,
Department of Industrial Engineering
Yıldız Technical University

Date Approved: 14 June 2021

To my family

ABSTRACT

The Plastic injection molding machine (IMM) has been one of the most preferred manufacturing methods since its invention and continues its development in parallel with the advancing technology. The demand for faster, better quality, and cheaper parts with improved quality output is increasing day by day. Manufacturers are dealing with quality errors in production while trying to meet this demand. The most common of these quality errors are the short shot missing injection error. The special mold investment, changing of part design, and usage of distinct plastic raw materials are carried out to solve this error. The most practical temporary action taken to prevent this error in production is to increase the raw material's weight by optimizing the process parameters. Unfortunately, all of these solution methods increase the product cost.

This thesis aims to develop a system for preventing short shot quality failure without using a solution to massive increase the part's cost. Firstly, the finite element analysis (FEA) for the short shot quality defect in injection machinery molding (IMM) is performed via Autodesk Moldflow Analysis. The mold flow FEA model is verified in serial production. Then, IMM is integrated into manufacturing execution system (MES) and a quality control system with cameras installed at the end of the production line. The relationship between short shot quality defect and IMM critical process parameters extracted with data analytic. The model is generated and commissioned in the production line. The integration of each production line into MES and installing a camera measurement system can adversely affect industrial deployment sustainability. Therefore, a different approach is developed with in-mold pressure/temperature sensors.

The production parameters affecting the short shot quality failure are determined via the data analytics. The goal is to produce the system in the ideal range in each cycle by giving feedback to the IMM according to the in-mold sensors' data during production. The main aim of this thesis is to avoid short-shot quality failure in plastic part production. The short-shot quality defect has not been eliminated in this study. However, it has decreased from 3.7% error rates to 0.82%. This dramatic reduction results in an overall 80% improvement in short shot quality failure without increasing product cost.

ÖZETÇE

Plastik enjeksiyon prosesi, ilk üretimde kullanıldığından bu yana en çok tercih edilen imalat yöntemlerinden biri olmuş ve gelişen teknolojiye paralel olarak gelişimini sürdürmektedir. Plastik parçalara olan daha hızlı, daha kaliteli ve daha ucuz talebi her geçen gün artmaktadır. Üreticiler bu talebi karşılamaya çalışırken üretimde kalite hataları ile uğraşmaktadırlar. Bu kalite hatalarından en yaygın olanı, kalıbın dolmaması olarak bilinen eksik enjeksiyon hatasıdır. Bu hatayı seri imalatta gidermek için özel kalıp yatırımı, parça tasarımının değiştirilmesi ve farklı plastik hammaddelerin kullanılması denenmektedir. Hata üretimde bu hatayı önlemek için alınan en pratik geçici eylem, proses parametrelerini optimize ederek hammadde ağırlığını artırmaktır. Ne yazık ki tüm bu çözüm yöntemleri ürün maliyetini artırmaktadır.

Bu tez, plastik parçanın maliyetini artırıcı bir çözüm kullanmadan eksik enjeksiyon kalitesi önlemek için bir sistem geliştirmeyi amaçlamaktadır. İlk olarak, potansiyel eksik enjeksiyon hatasının bölgelerinin tespit edilemesi için sonlu elemanlar yöntemi kullanılarak kalıp akış analizi yapan Autodesk Moldflow programı kullanılmıştır. Analiz sonuçları üretimde denenerek doğrulanmıştır. Ardından plastik enjeksiyon makinesi ve hat sonuna kurulan kamera kontrol sistemi üretim yönetim sistemine entegre edilmiştir. Kritik proses parametreleri ile eksik enjeksiyon kalite çıktısı veri madenciliği ile istatistisel olarak ilişkilendirilip, seri üretimde devreye alınmıştır. Her üretim hattının sonuna kamera kurmak, plastik enjeksiyon makineleri üretim yönetim sistemine almak endüstriyel yaygınlaştırmayı olumsuz etkileyebileceğinden kalıp içi basınç/sıcaklık sensörleri ile farklı bir yaklaşım getirilmiştir.

Data analitik sonucunda bulunan eksik enjeksiyona etki eden üretim parametreleri belirlenmiştir. Bu kapsamda üretim sırasında kalıp içi sensörlerin verilerine göre plastik enjeksiyon makinasına geri bildirim vererek sistemi her çevrimde ideal aralıkta üretmek amaçlanmıştır. Yapılan çalışmalarda eksik enjeksiyon hatası tamamen giderilememiştir. Ancak ilk durumdaki %3.7 rework oranından %0.82 seviyelerine düşürülmüştür. Bu azalma ürün maliyetini artırmadan eksik enjeksiyon hatasını %80 iyileştirmiştir.

ACKNOWLEDGMENTS

I want to express the most profound appreciation to my thesis advisor Assistant Professor Erinç Albey, for his guidance, patience, and supports. His dedication to the job, diligence, and research motivation has always impressed me.

I am also indebted to Arçelik for providing the research tools, which my entire research was based on and special thanks to my colleagues that are Vügar Kerimoğlu, Gökay Gökmen, Ömer Faruk Özer and Gökhan Engin for sharing this valuable experience with me.

Thanks to Keyence and Priamus companies for their agile approach and support in the process.

Finally, I am most thankful to my wife for their unlimited support throughout my entire life. Nothing would be possible without her sacrifice and patience. I dedicate this work to her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

An injection molding machine could be a producing process that has to form the plastic raw material melted with temperature by injecting it into a mold and taking it from the mold by cooling. With this process, it's conceivable to form many materials of various sizes. The plastic injection strategy is one of the foremost common production processes utilized in the industry. The machine within which the plastic manufacturing process is administered is named a plastic injection machine. It consists of three main units: the molding unit, clamping unit, and injection unit. The section, called the Clamping Unit, serves to keep the mold under pressure during the cooling and injection processes. It provides to join the two sides of the injection mold.

1.1 Literature Survey

During the injection process, the granulated plastic material is poured from the chamber located above the injection unit. it's then taken into a cylinder (barrel), which is heated by heaters with resistance on it, with the assistance of a screw controlled by an electrical motor. The screw advances the molten plastic material to the end of the cylinder by being tightened continuously. Then the injection process is started. The manufacturing process of a plastic injection part consists of the subsequent stages;

- Mold Closing: An injection molding machine comprises three primary parts: clamping unit, injection unit, and mold unit. The clamping unit, injection, and cooling application is that the unit under pressure. More basically combining two sides (female and male) of a plastic injection mold.

- Plasticization (Melting): The barrel, screw, and raw material are heated by the heating elements at the bucket's level. The raw material is compressed between the pitches while the screw rotates, rubbing and producing heat. The raw material standing on the spinning screw slowly melts into powder, granules, soft melt, high viscosity melt, or semi-hard powder. At the end of the screw passes, the rocket group starts to accumulate in front of the screw. The rocket group begins to gather in front of the screw as the screw moves. The hydro motor spins the screw at the desired rpm, transfers the raw material in front of it like a revolving screw conveyor, and fills it. The charged plastic becomes twisted and starts to pull the screw out. The oil is compressed with the back control mechanism when it returns from the injection cylinder so that the screw does not go out too far and there are no air pockets in the volume of counter-pressure from the finishing operation.
- Injection: As the screw rotates backward, the bracelet slides and rests against the rocket. The ferrule and the collar are handled in front of the screw that can be applied. When the previously programmed machine detects enough plastic to fill the mold with micro switches or transducers, it stops operations, turning the screw and going back. This time the injection cylinder pushes the screw forward with a programmed speed and pressure. The ring leaks back and presses high, and prevents the material from escaping back like a check valve. The melting plastic passes through the nozzle and runner and fills the mold.
- Holding: It presses the screw for a set amount of time to prevent the part from collapsing and shrinking due to cooling in the mold, and the part is sized. Through applying friction, the plastic melt inserted into the mold is packed with pressure. The procedure is repeated until the plastic inside the mold has solidified. The

duration of the process varies depending on the characteristics of the raw material, part size, and weight.

- Cooling: As the timer runs out, the pressure is reduced. The plastic substance inside the mold has been frozen, and the mold has been cooled. The screw begins to spin again after taking advantage of the freezing moment. When returning, the pressing of the plastic begins. When the plastic has been removed, the screw is retracted 5-10 mm as required, and the material in the nozzle mouth is sucked out. The products in the nozzle are separated from the plastic in the runner.
- Mold Opening: The formed substance is removed from the female mold as the mold is lifted, and there is no runner break. The nozzle fluid does not leak immediately when the runner is empty, and it does not clog the runner or flow into the mold. The nozzle and the top of the bucket are heated in a managed manner during this operation.
- Part Removal: The pushers are used to separate the finished part from the mold.

Injection molding machines can be configured in a variety of ways, with a variety of components (single or multiple units), which can be horizontal or vertical. Injection molding devices, on the other hand, all follow the same general procedure. The procedure outlined above is applicable to every injection molding system.

Figure 1: Injection Molding Machine

Plastic injection molded components, like any other production procedure, may have complications and failures that require troubleshooting. The short shot quality failure, which is caused by the mold not filling, is one of the most popular IMM quality issues. The mold stops before entirely filling the mold cavity, resulting in holes or thin areas on the component surface and a faulty product. These holes are commonly located in the mold's thinnest regions or on the edges of the finished product. In a traditional plastic injection process, these are the last positions a mold would fill.

As one of the most common plastic injection molding problems, there are many reasons why short shots occur, but they fall into three main categories:[1]

- Problems with the plastic selection.
- Problems with mold design.
- Problems with the IMM.

Potential Short Shot Failure Due to Material Selection

The acceleration of melted plastic flow through the mold cavity is a recurring theme in most of the issues. The most likely cause is something that interrupts this flow and causes the mold to fill insufficiently. The selected plastic should have a high viscosity for use if the mold flow is not analyzed. The Melt Flow Index is normally used to verify it (MFI). Flow analysis can be used to choose the plastic with the required viscosity value based on the usage condition.[2]

Potential Short Shot Failure Due to Mold Design

One of the most important aspects of the plastic injection process is the mold. Short shot quality defects may be caused by a poorly built mold. [4]

Potential Short Shot Failure with the Molding Machine

It's crucial to match the right component to the right molding machine for successful manufacturing and to prevent flow and pressure problems. If the shot size is less than 25% of the machine's power, a small machine may not have enough pressure or ram speed, whereas a large machine can have trouble handling pressure. Higher temperatures in the barrel and die wall will also aid in the rapid cooling of molded plastic. Understanding the various models of plastic injection molding machines and how different types and sizes will gain is critical. [3]

1.2 Objective

According to the literature, the plastic injection procedure should be designed. It begins with selecting plastics with acceptable flow characteristics based on MFI values, a component design that will not trigger melt plastic flow issues, and machinery that meets the necessary specifications. This method allows ramp-up output to begin after the procedure has been developed. With the successful completion of the ramp-up phase, serial production will begin.

Despite this, due to process instability, short shot output loss occurs during manufacturing. To correct this error, the process parameters are changed by optimizing process settings. It gives a partial solution and recurs after a certain amount of time has passed. To overcome this recurring mistake, unique component designs, dual plastic components, and special mold manufacturing have been attempted. Regrettably, both of these solutions raise the price of the commodity.

The aim of this study is to prevent short shot quality loss without resorting to a solution that dramatically increases the part's cost.

1.3 Layout of the Thesis

The organization of this thesis is as follows:

In Chapter 1, the plastic injection process is explained along with its steps. Causes of short shot failure are examined. It is stated how solutions can be obtained with the material used, cost-increasing part design, and special mold manufacturing. The purpose of the thesis is explained.

In Chapter 2, the FE model of the mold flow analysis is created. The mold flow analysis results are interpreted, and potential short shot quality failure areas are evaluated.

Chapter 3 discusses the work on integrating the IMM machine into the MES infrastructure. It refers to the integration of the system with the OPC-UA communication protocol. The camera control system installed at the end of the production line is explained. The verification of potential short shot quality failure areas compared as a result of FEA with the camera control system.

In Chapter 4, the critical process parameters are collected from the IMM machine by data acquisition. The relationship between process parameters and the camera control system is statistically modeled via data mining. The predictive quality model is and implemented in serial production. It is explained that in-mold sensors are used for industrial deployment by increasing the metric performance successes of the model. The results of these studies are shown and compared to the current situation with results.

In Chapters 5 and 6, suggestions for the further research and discussions are presented.

CHAPTER 2

MOLD FLOW FEA

This section includes fluid flow approach, assumptions, and mold flow analysis for brief short shot quality failure. Mold flow analyses are performed to determine the behavior of critical parts within the plastic injection cycle. This analysis simulates a plastic injection molding cycle with a specific plastic and analyzes the results. In the chapter, firstly, the literature researches about the mold flow physical model are explained. The mathematical modeling approaches are introduced by Zhou , Kumar and, Courbebaisse. Then assumptions for the washing machine tub part are explained. The finally, Autodesk Moldflow adviser fill and pack analysis results are shared.

2.1 Fluid Flow Governing Equations

The mass conservation equation, momentum conservation equation, and energy conservation equation are the governing equations for plastic injection melted flow. The standard governing equations can be made more widely used by using these three equations, which have the same structure. These numbers, which govern the flow of compressive fluids, are shown to function in the polymer melting phase and are calculated using density, momentum, and energy conservation principles. [5]

The equations of governing the behavior of a fluid motion are explained by Zhou [5]. The finally which are the continuity, momentum, and energy general equations shown as:

$$\frac{\partial \rho}{\partial t} + \nabla.(\rho u) = 0 \quad (1)$$

$$\frac{\partial (\rho u)}{\partial t} + \nabla.(\rho u u) = \nabla.(\sigma) + \rho f \quad (2)$$

$$\rho c \frac{\partial T}{\partial t} + c\nabla.(\rho T u) = \nabla.(k.\nabla T) + \tau_{xx}\frac{\partial u}{\partial x} + \tau_{yx}\frac{\partial u}{\partial y} + \tau_{zx}\frac{\partial u}{\partial z} + \tau_{xy}\frac{\partial v}{\partial x} + \tau_{yy}\frac{\partial v}{\partial y} + \tau_{zy}\frac{\partial v}{\partial z} + \tau_{zx}\frac{\partial \omega}{\partial x} + \tau_{zy}\frac{\partial \omega}{\partial y} + \tau_{zz}\frac{\partial \omega}{\partial z} + S_i \quad (3)$$

It is not feasible to resolve industrial numerical techniques software in complicated domains like injection mold cavities with equations approach.

2.1.1 Hele-Shaw Model

The filling, packaging, and cooling are the three main phases in the injection process. The polymer melt is injected into a cold freezing hole in the filling process, where it expands under high pressure and fills the mold. High pressure is retained after the mold is filled in the packing point, and additional melt flows into the cavity to compensate for density changes during cooling. The melt is then cooled, and the formed product is ejected in the final stage. [6]

Several assumptions are made to make material data requirements more reasonable. The Hele-Shaw model is the most commonly used model for plastic injection manufacturing simulation. It is applicable to flows in the narrow gap. Conventionally, most plastic products have thin shell structures**.** As a result, the filling flow in a very thin cavity can be approximated by Hele-Shaw flows:

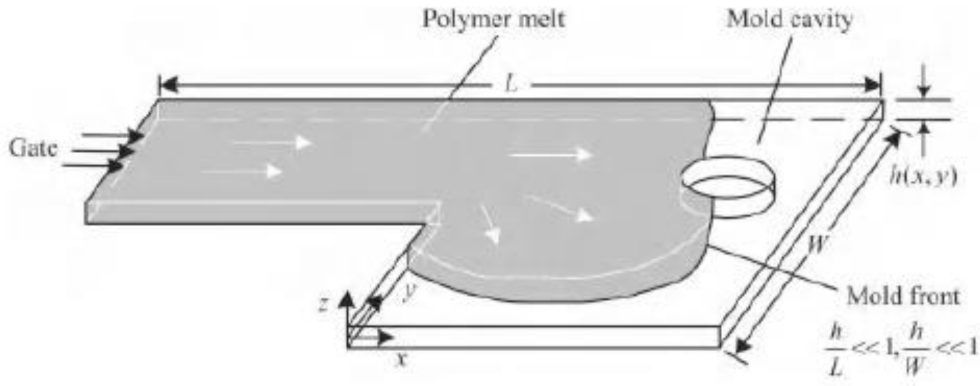

Figure 2: A schematic diagram of a typical flow described by the Hele-Shaw model (Source: Zhou[5])

The following assumptions guide this method's approximation approach:

1. Consider melting flow in an incompressible fluid's filling phase newtonian fluid. The pressure follows an uniform approach, and the momentum in depth can be ignored.
2. Neglect the inertial force, which is much less than the viscous shear stress due to the high viscosity of the fluid.
3. Heat convection in the width direction is ignored.
4. Neglect heat conduction because heat convection is a slightly more significant issue.
5. Inability to recognize turbulence in the melting plastic flow rate.
6. To neglect width direction on simulation.

The Hele-Shaw (Shell model) is an ideal approach to simulate this kind of tub part in IMM.

2.1.2 Filling Phase Assumption

The melt can be incompressible during the filling process. This assumption concerning of the this approach included the constant density and the surface dilation viscosity λ=0.

$$\nabla . u = 0 \quad (4)$$

The melt can reflect a Generalized Newtonian Fluid during the filling process. Under this assumption, viscoelastic effects can be ignored. The fundamental interaction can be used for generalized newtonian fluid viscosity models.

$$\frac{\partial(\rho u)}{\partial t} + \rho(u . \nabla u) = -\nabla \rho + \nabla \eta(\nabla u + \nabla u^T) + \rho f \quad (5)$$

In the IMM filling phase, the plastic melt's thermal conductivity is assumed to be constant. Despite the fact that the thermal conductivity k of polymers is temperature dependent, this assertion is used due to the difficulties in collecting material details.

$$\frac{\partial(\rho i)}{\partial t} + \nabla . (\rho i u) = k \nabla^2 T + \Phi + S_i \quad (6)$$

2.2 Moldflow Analysis

Moldflow plastic insight (MPI) is a software program that allows you to analyze plastic parts and molds. The aim of using such an inspection software is to predict potential surface defects in order to produce a high-quality product. MPI Software simulates the mold in the manufacturing process to determine mold design and parameter settings. This type of advancement will greatly boost injection mold design performance while lowering production costs. [8] The analysis process is shown in Fig 3.

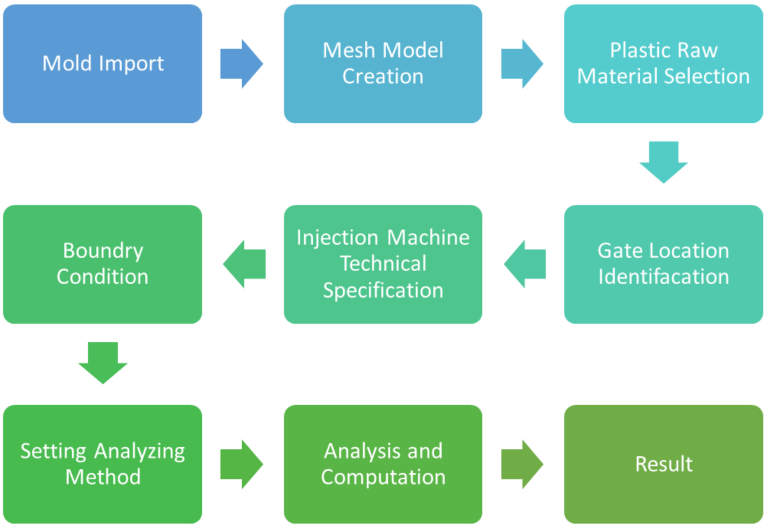

Figure 3: The rheological analysis process

2.2.1 Autodesk Moldflow Analysis Pre-Process

The assumption that the mesh components are uniform, equilateral tetras underpins the mathematical study of a model. Mesh tetras thus vary from the same volume of tetra to distortion that has a detrimental impact on the data. As a result, it's important to cut down on the number of twisted tetras as much as possible.

The most common method for calculating mesh efficiency is the Aspect Ratio. Low aspect ratio elements are needed in mesh-sensitive areas such as gates and major geometry adjustments. The optimal tetra with similar edge lengths has an aspect ratio of 1 according to the FEA program. The rate of a flat tetra is infinite. The objective is to keep the overall aspect ratio below 30. For such applications, a maximum aspect ratio of 50 is recommended.

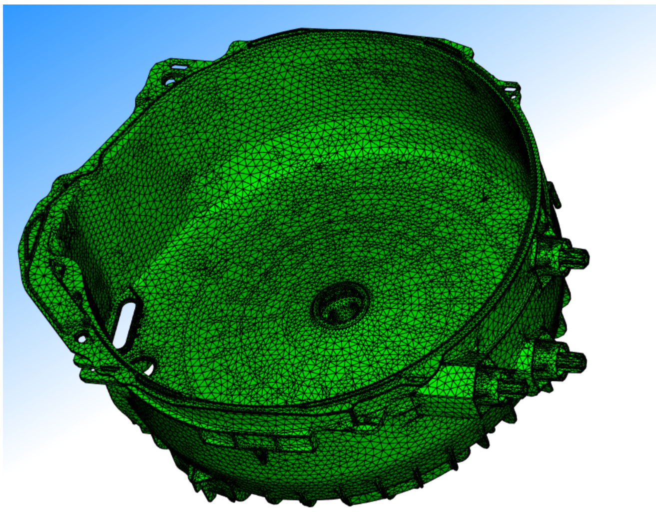

Figure 4: Mesh Structure

During pre-processing, the aspect ratio is tested on a regular basis. The only way to control tetrahedral mesh systems is to intervene in key design areas. Throughout the process, the smooth mesh feature turned on. Surface aspect ratio: Maximum = 14.79 average = 2.78

Table 1: Technical Specification

Maximum pressure:	180	MPa
Screw intensification ratio:	10	-
Machine maximum clamp force:	7000	Ton
Mold temperature:	20	C
Transition temperature:	116	C
Melt temperature:	220	C
Trade Name:	Polyarr	-
Family Name:	PP	-

2.2.2 Autodesk Moldflow Analysis Post-Process

Autodesk Moldflow's fill and pack analysis examines the polymer flow within the mold during the filling and packing processes. These studies are used to see if a cavity is filled up. The pack analysis starts after the fill analysis. To predict polymer reaction, the pack analysis uses several properties calculated during the filling process. Figure 4 shows the contour colors reflecting the movement of plastic through the component based on the filling time effect. The same color covers all of the regions at the same time. The outcome is dark blue at the beginning of the injection and red in the final regions. The unfilled portion of a short shot has no colour. A potential short shot consistency loss due to process instability can be seen in red in the filling areas.

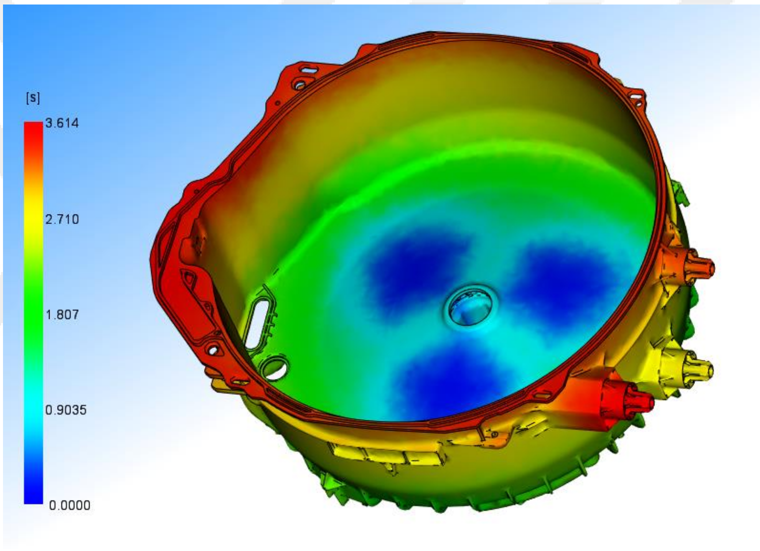

Figure 4: Fill Time Result

The V/P switch-over determines the proportion of the total volume where the injection molding machine (IMM) switches from filling to packaging. Determines whether the process control logic is velocity-driven (fill) and pressure-driven (pack). As a result, places that aren't filled within the V/P transformation are grayed out. Figure 5 represents the pressure value at the V/P switchover stage.

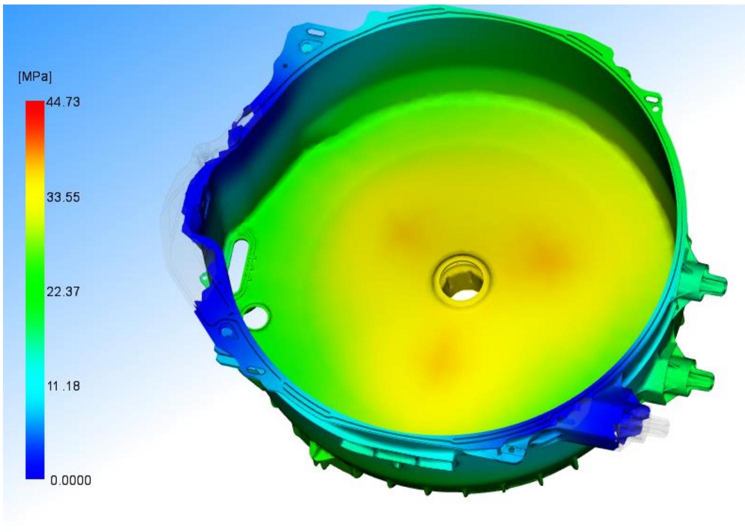

Figure 5: Pressure at V/P switchover

The temperature in the previous flow effect, which are produced by the filling measurement, refers to the temperature of the polymer when the flow front reaches the stated position within the plastic cross's center. The temperature at the flow front result uses a range of colors to point from the lowest temperature in blue to the highest temperature in red, as seen in the diagram below. The colors reflect the temperature of the material at each point. The temperature of the flow front varies during filling, as seen in the result. More significant temperature variations sometimes mean that the injection period is simply too short or that there are areas of hesitation. If the flow front temperature is too low in a particular region, uncertainty can vanish in an instant. Surface defects may be caused by a sudden change in pre-flow temperature. Figure 6 shows the sections of which special attention should be given in blue to the color contours. The local temperature deteriorates, potentially resulting in poor short-shot areas.

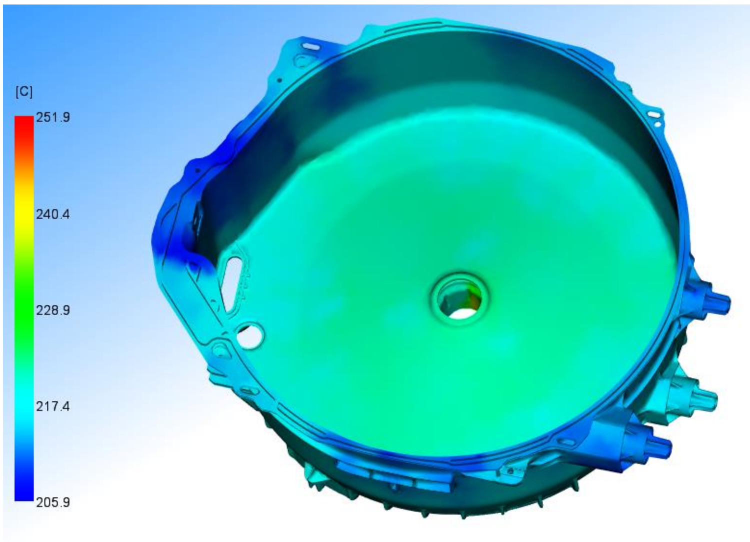

Figure 6: Temperature at the flow front

A filling analysis result reveals polymer melt behaviour in the pressure distribution at the tip of the filling step. The pressure in the mold at the start of filling is 1 atm on the absolute pressure scale. Only when the melt front approaches a certain location does the pressure at that location begin to increase. The force driving the polymer dissolves the flow during filling due to the change in friction from one stage to another. The pressure differential is the difference in pressure between two points separated by the distance between them. Polymer always moves in the negative pressure gradient direction, from high to low pressure.

As a result, at the filling point, the highest pressure occurs at the polymer injection positions and the lowest pressure occurs at the melt front, as seen in Figure 7. The large differences in pressure distribution, as shown by neighboring lines, can be avoided during the filling process. Variations in pressure during packaging have an effect on volumetric shrinkage. Also at the packaging period, pressure variations inside the cavity should be kept to a minimum.

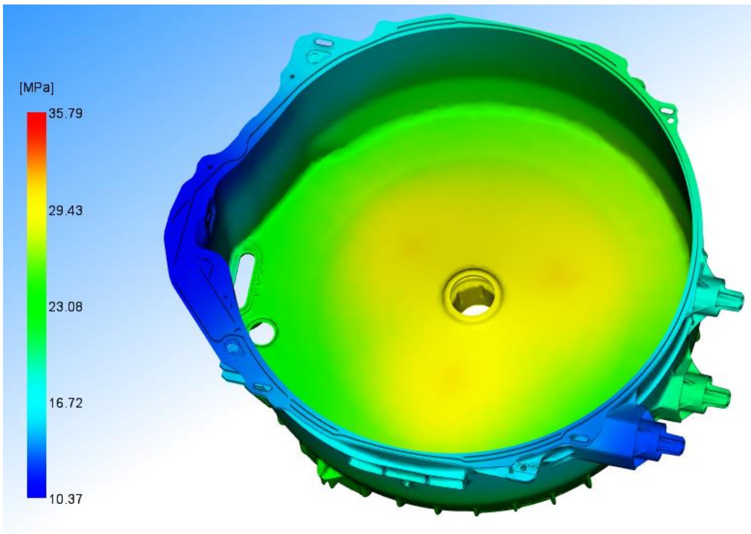

Figure 7: Pressure at end of fill

2.3 Comparison of Moldflow Analysis with Actual Production Results

The improvements are made to avoid potential errors partly design via mold flow analysis results. However, despite these improvements, the short shot problem has been observed due to the process variability.

Figure 8: Short Shot Quality Failure in Production

Moldflow research reveals potential short-shot quality defect areas that coincide with manufacturing defects. Taking into account on the process change, the error seen in Figure 8 happens. As a result, experiments of process parameters are conducted, but no definitive optimization has been achieved. Output has been programmed with process parameters in order to solve this mistake. However, no transparent solution has been discovered. Using the plastic injection production's experience, the typical loss is attempted to be avoided. Inexperienced technicians, on the other hand, are more likely to make short shot mistakes in serial processing.

2.4 Summary

Moldflow research reveals potential short-shot quality defect areas that coincide with manufacturing defects. Taking into account on the process change, the error seen in Figure 8 happens. As a result, experiments of process parameters are conducted, but no definitive optimization has been achieved. Output has been programmed with process parameters in order to solve this mistake. However, no transparent solution has been discovered. Using the plastic injection production's experience, the typical loss is attempted to be avoided. Inexperienced technicians, on the other hand, are more likely to make short shot mistakes in serial processing.

In this section, the injection machine molding (IMM) mold flow melt plastic flow model utilized in this study is explained very well. The governing equations for plastic injection melted flow have clarified the mass conservation equation, momentum conservation equation, and energy conservation equation. From these equations, the Hele-Shaw model is formed with its assumptions concerning the literature review. Then, the flow is calculated with this solution model using FEA software. Autodesk Moldflow, which is employed industrially, is chosen for this application.

As a results of the analysis, potential short shot quality failure regions are extracted. The study is verified by comparing them with the failure zones observed in serial production trials. This study aims to detect the error with cameras installed in potential areas when the standard failure occurs counting on process variability.

CHAPTER 3

DATA INFASTRUCTURE

In this section, the concept of the digital twin introduces by Jones, Lothander, Zhuang. Within the scope of data acquisition, the OPC-UA communication protocol is mentioned. It investigated whether MES integrated into the IMM machine. How to establish a technical architecture for data-driven models introduces by Teng and Gray. In the previous section, potential short shot quality defect areas determined by mold flow analysis are tried to be controlled by camera setup. With Popp, Heleno, Mohan, and Amza, it is determined how a system infrastructure is created. The processing of the logged photos with image processing and their commissioning in production are explained.

3.1 Data Acquisition

The first step is to create the virtual representation of physics of the IMM process; in other words, the digital twin. Consisting of a physical product, its virtual model, and the bi-directional data connections, a digital twin, is a tool for enhancing physical entities' performance by leveraging computational techniques. [9] Digital twins also align with the product life-cycle. [10] By virtual-to-physical information flow, a digital twin contains the functionality to physically realize a change in the physical state as practiced in process control [11], machine parameters [12], and production management. [13] Gartner estimates that by 2021, half of all large industrial enterprises will use a digital twin, and those users will become 10% more effective [14].

Developing a digital twin by simulating injection machinery molding (IMM) physics requires experience in the plastic injection process and complex simulation expertise. To understand the complex behavior of a production process to improve quality and estimate the material volume needed for process modeling. The project has consulted with experienced operators and production engineers and analyzed how operators control the process and how operators overcome the quality problems. This way, it could maintain user involvement and sustain the continuous information flow from the real environment.

3.1.1 Communication Protocol

OPC Unified Architecture (OPC UA) is an industry communication platform that is interactive. The OPC Foundation designed the platform. The service-oriented architecture is included in the OPC UA (SOA). [15] OPC Basic Services is a system that does not depend on the protocol to operate. [16]

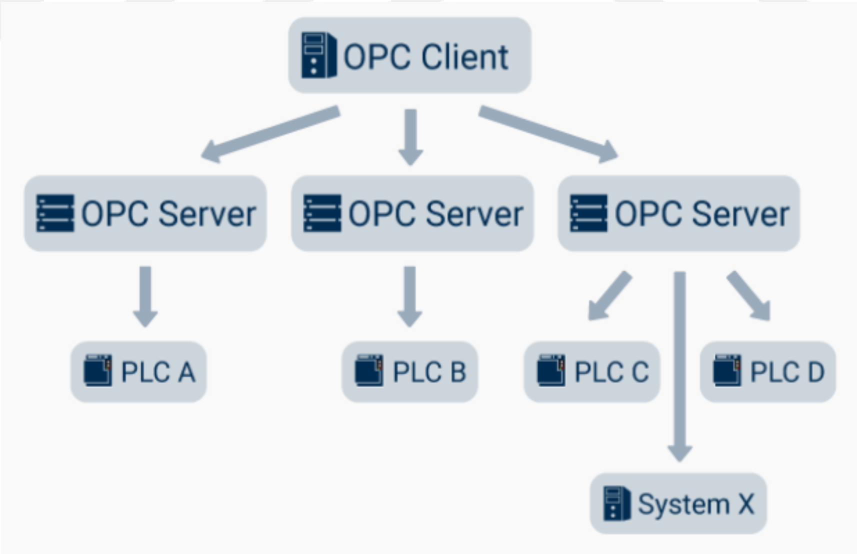

Figure 9: OPC Server Architecture (Source OPC UA 2019)

The system's communication form is the OPC Server. It is software that implements the OPC specification to have standardized interfaces. OPC makes use of a special collaboration protocol. The structure of OPC Servers can be conveniently programmed, and they are controlled in various areas. Industrial automation is developed as an interface that provide standard read and write access when conventional OPC standards are first developed.

EUROMAP 77 establishes a connection between injection molding machines (IMMs) and production management information systems (MES). MES stores data generated by IMM on central servers to make factory management easier. It provides an interface for IMM and MES of various machine manufacturers when used in connection with EUROMAP 77.

3.1.2 Technical Architecture

The sensor, experimental measurements, and domain expertise techniques are the most popular models used in industry. [17] Furthermore, the real problem of Industry 4.0 is ensuring that the technology is well set up and that the system continues to operate in a sustainable manner. One of the most pressing issues is data consistency. [18]

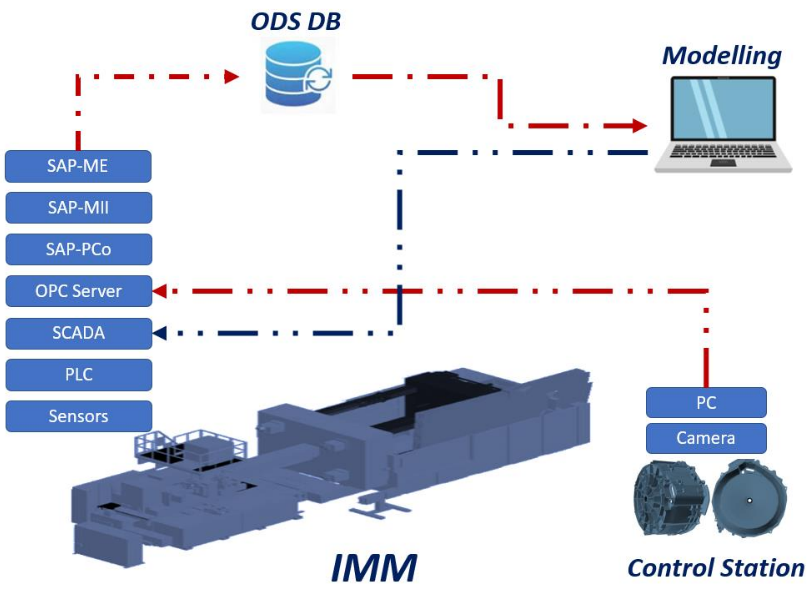

Figure 10: Technical Architecture Diagram

The OPC Server collects machine parameters using the OPC UA protocol and the IMM EUROMAP 77 interface. The short shot mistake is photographed using the camera quality management device that will be mounted at the end of the line. It is given in order for quality information to be logged to the OPC Server. With MES, this labelled output data is saved to ODS DB. The data is analyzed by the creator as they create a web service. The consistency prediction models have been checked to run on IMM SCADA. The consistency predictive model will be possible in the future AWS depending on infrastructure improvements.

3.1.2 Web Service Creation

A web server is a collection of protocols and standards for exchanging data between programs or systems. Web platforms can also be used for programs written in various programming languages to share data over computer networks such as the internet. The web services are created to access ODS DB. It can be accessed on the ME/MII server.


```
<?xml version="1.0" encoding="UTF-8"?>
<Rowsets Version="15.2 SP0 Patch 8 (Jun 29, 2018)" StartDate="2019-03-12T12:36:54" EndDate="2019-03-12T13:36:54" DateCreated="2019-03-12T13:37:00" CachedTime="">
 - <Rowset>
   - <Columns>
       <Column SourceColumn="Resource" SQLDataType="-9" Name="Resource" MinRange="0" MaxRange="1" Description="Resource"/>
       <Column SourceColumn="Operation" SQLDataType="-9" Name="Operation" MinRange="0" MaxRange="1" Description="Operation"/>
       <Column SourceColumn="Item" SQLDataType="-9" Name="Item" MinRange="0" MaxRange="1" Description="Item"/>
       <Column SourceColumn="SFC" SQLDataType="-9" Name="SFC" MinRange="0" MaxRange="1" Description="SFC"/>
       <Column SourceColumn="Datetime" SQLDataType="93" Name="Datetime" MinRange="0" MaxRange="1" Description="Datetime"/>
       <Column SourceColumn="Measures" SQLDataType="-9" Name="Measures" MinRange="0" MaxRange="1" Description="Measures"/>
     </Columns>
   - <Row>
       <Resource>I022000</Resource>
       <Operation>I010</Operation>
       <Item>1910500100,A</Item>
       <SFC>2511031900003876</SFC>
       <Datetime>2019-03-11T08:25:38</Datetime>
       <Measures>[{"MeasureName":"Carriage backward time","ActualValue":"0.0"},{"MeasureName":"Carriage forward time","ActualValue":"0.0"},{"MeasureName":"Clamp close time","ActualValue":"7.59999999999939"},
         {"MeasureName":"Clamp force","ActualValue":"13044.6449497768"},{"MeasureName":"Clamp open time","ActualValue":"8.23799999999942"},{"MeasureName":"Cooling time","ActualValue":"29.9999999999943"},
         {"MeasureName":"Cooling ring temperature","ActualValue":"0.0"},{"MeasureName":"Cycle time","ActualValue":"0.0"},{"MeasureName":"Decompress time","ActualValue":"3.04399999999989"},{"MeasureName":"Ejector
         backward time","ActualValue":"1.4e-2"},{"MeasureName":"Ejector forward time","ActualValue":"2.31599999999997"},{"MeasureName":"Holding Position","ActualValue":"44.9243998358526"},{"MeasureName":"Holding
         time","ActualValue":"1.004"},{"MeasureName":"HotRunner Zone 1 temperature","ActualValue":"250.2"},{"MeasureName":"HotRunner Zone 10 temperature","ActualValue":"249.9"},{"MeasureName":"HotRunner Zone 11
         temperature","ActualValue":"249.9"},{"MeasureName":"HotRunner Zone 12 temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone 13 temperature","ActualValue":"250.3"},{"MeasureName":"HotRunner Zone
         14 temperature","ActualValue":"249.9"},{"MeasureName":"HotRunner Zone 15 temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone 16 temperature","ActualValue":"250.3"},{"MeasureName":"HotRunner
         Zone 17 temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone 18 temperature","ActualValue":"250.1"},{"MeasureName":"HotRunner Zone 19 temperature","ActualValue":"250.0"},
         {"MeasureName":"HotRunner Zone 2 temperature","ActualValue":"249.9"},{"MeasureName":"HotRunner Zone 20 temperature","ActualValue":"250.4"},{"MeasureName":"HotRunner Zone 21
         temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone 22 temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone 23 temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone
         24 temperature","ActualValue":"250.1"},{"MeasureName":"HotRunner Zone 3 temperature","ActualValue":"249.8"},{"MeasureName":"HotRunner Zone 4 temperature","ActualValue":"250.2"},{"MeasureName":"HotRunner Zone
         5 temperature","ActualValue":"250.0"},{"MeasureName":"HotRunner Zone 6 temperature","ActualValue":"250.2"},{"MeasureName":"HotRunner Zone 7 temperature","ActualValue":"249.9"},{"MeasureName":"HotRunner Zone 8
         temperature","ActualValue":"250.2"},{"MeasureName":"HotRunner Zone 9 temperature","ActualValue":"250.1"},{"MeasureName":"Injection end position","ActualValue":"30.9676595308119"},{"MeasureName":"Injection
         position 1","ActualValue":"1.0"},{"MeasureName":"Injection position 10","ActualValue":"415.0"},{"MeasureName":"Injection position 2","ActualValue":"20.0"},{"MeasureName":"Injection position 3","ActualValue":"50.0"},
         {"MeasureName":"Injection position 4","ActualValue":"150.0"},{"MeasureName":"Injection position 5","ActualValue":"180.0"},{"MeasureName":"Injection position 6","ActualValue":"180.0"},{"MeasureName":"Injection position
         7","ActualValue":"180.0"},{"MeasureName":"Injection position 8","ActualValue":"180.0"},{"MeasureName":"Injection position 9","ActualValue":"180.0"},{"MeasureName":"Injection pressure","ActualValue":"0.0"},
         {"MeasureName":"Injection pressure 1","ActualValue":"50.0"},{"MeasureName":"Injection pressure 10","ActualValue":"175.0"},{"MeasureName":"Injection pressure 2","ActualValue":"70.0"},{"MeasureName":"Injection
         pressure 3","ActualValue":"100.0"},{"MeasureName":"Injection pressure 4","ActualValue":"130.0"},{"MeasureName":"Injection pressure 5","ActualValue":"100.0"},{"MeasureName":"Injection pressure
         6","ActualValue":"120.0"},{"MeasureName":"Injection pressure 7","ActualValue":"125.0"},{"MeasureName":"Injection pressure 8","ActualValue":"0.0"},{"MeasureName":"Injection pressure 9","ActualValue":"120.0"},
         {"MeasureName":"Injection segment 1 inuse","ActualValue":"TRUE"},{"MeasureName":"Injection segment 10 inuse","ActualValue":"TRUE"},{"MeasureName":"Injection segment 2 inuse","ActualValue":"TRUE"},
         {"MeasureName":"Injection segment 3 inuse","ActualValue":"TRUE"},{"MeasureName":"Injection segment 4 inuse","ActualValue":"TRUE"},{"MeasureName":"Injection segment 5 inuse","ActualValue":"FALSE"},
         {"MeasureName":"Injection segment 6 inuse","ActualValue":"FALSE"},{"MeasureName":"Injection segment 7 inuse","ActualValue":"FALSE"},{"MeasureName":"Injection segment 8 inuse","ActualValue":"FALSE"},
         {"MeasureName":"Injection segment 9 inuse","ActualValue":"FALSE"},{"MeasureName":"Injection Speed","ActualValue":"56.6070207236166"},{"MeasureName":"Injection switchover
         position","ActualValue":"44.9243998358526"},{"MeasureName":"Injection time","ActualValue":"7.4899999999994"},{"MeasureName":"Injection velocity 1","ActualValue":"11.8125"},{"MeasureName":"Injection velocity
         10","ActualValue":"78.75"},{"MeasureName":"Injection velocity 2","ActualValue":"47.25"},{"MeasureName":"Injection velocity 3","ActualValue":"55.125"},{"MeasureName":"Injection velocity 4","ActualValue":"63.0"},
         {"MeasureName":"Injection velocity 5","ActualValue":"0.0"},{"MeasureName":"Injection velocity 6","ActualValue":"0.0"},{"MeasureName":"Injection velocity 7","ActualValue":"0.0"},{"MeasureName":"Injection velocity
         8","ActualValue":"0.0"},{"MeasureName":"Injection velocity 9","ActualValue":"0.0"},{"MeasureName":"Max injction speed","ActualValue":"56.6070207236166"},{"MeasureName":"Max open
         position","ActualValue":"2006.31172839506"},{"MeasureName":"Metering time","ActualValue":"23.5259999999979"},{"MeasureName":"Nozzle temperature","ActualValue":"234.1"},{"MeasureName":"Oil
         temperature","ActualValue":"22.0"},{"MeasureName":"Plastizing volume","ActualValue":"419.154784214486"},{"MeasureName":"Recipe Name","ActualValue":"E450A TUB"},{"MeasureName":"Shot
         Volume","ActualValue":"388.187124683674"},{"MeasureName":"Zone 1 temperature","ActualValue":"230.5"},{"MeasureName":"Zone 2 temperature","ActualValue":"235.4"},{"MeasureName":"Zone 3
         temperature","ActualValue":"233.6"},{"MeasureName":"Zone 4 temperature","ActualValue":"234.3"},{"MeasureName":"Zone 5 temperature","ActualValue":"234.5"},{"MeasureName":"Zone 6
         temperature","ActualValue":"229.3"},{"MeasureName":"Zone 7 temperature","ActualValue":"225.1"},{"MeasureName":"Zone 8 temperature","ActualValue":"217.9"},{"MeasureName":"Zone 9 temperature","ActualValue":"0.0"}]
         </Measures>
     </Row>
```


Figure 11: Response from Server to Client

3.2 Quality Control System

In the previous section, the critical regions for short shot quality failure are determined by mold flow analysis. The results of these analyzes have been verified by comparing them with the quality failure in serial production.

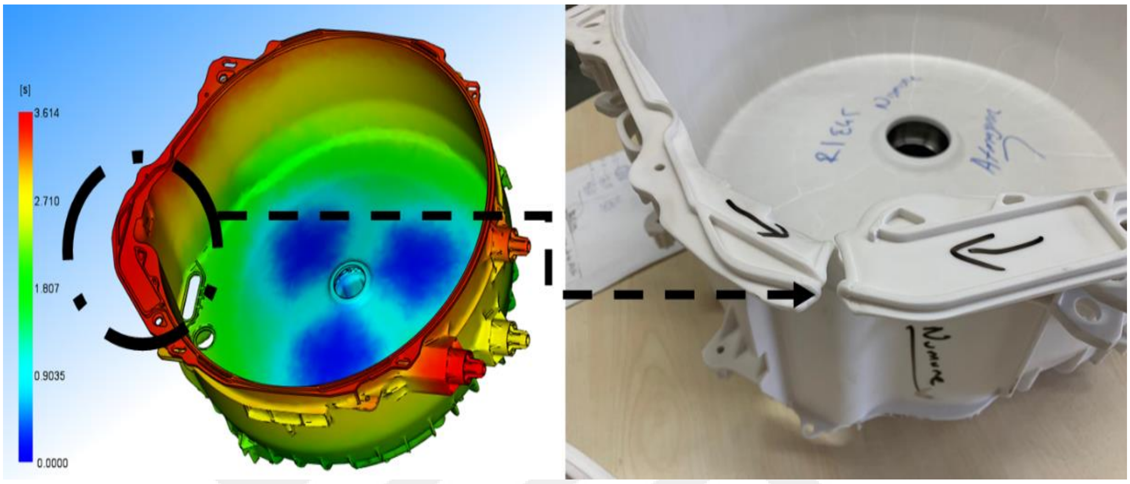

Figure 12: Short shot quality failure areas

The tub part of the washing machine is an important component in the product group, which includes the main features. Therefore, this error needs to be controlled in production. In the current situation, an operator's visual control is made at the end of the line. Depending on the operator's experience, if the error can be caught early, the machine can be intervened to prevent the error before it reaches a critical size. However, since the control system contains human factors, errors may occur serially in production depending on the operator's experience. In this case, the products are reworked. Firstly, it is considered to establish a quality control system to detect this error automatically.

3.2.1 Commissioning of the Quality Control System Equipment

In the industrial sector, quality management refers to the process of defining output requirements under optimal conditions and maintaining supply continuity. The quality management infrastructure must be in good shape for the manufacturing conditions to continue to run smoothly. Computer vision technologies are indispensable components of manufacturing processes due to their advantages and scalable structure in quality management. [19] In industrial manufacturing, vision systems play an important role in quality management systems. Such devices can now be installed easily, and research methods can be carried out at a low cost. [20]

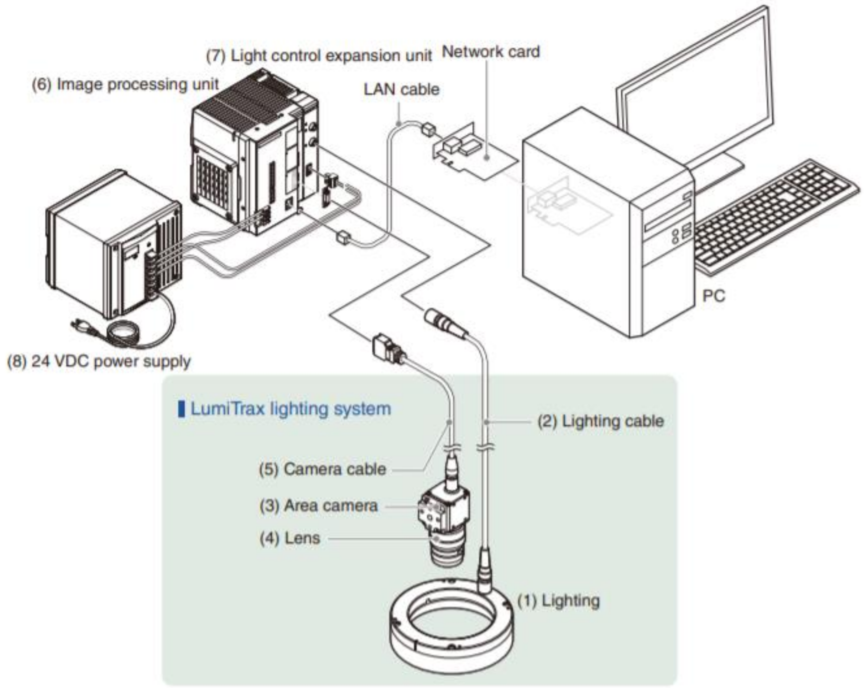

Figure 13: System Configuration Diagram (Source:Keyence)

In this application, a camera, lighting and image processing unit from the Keyence VJ series is selected. The system could set the shutter time and lighting control for the camera automatically. The image processing unit is device stack compliant. The VJ series can be applied to an existing system without requiring any changes to the infrastructure. The image processing unit is where the work is performed, and it can be programmed using OpenCV.

Figure 14: Quality Control System Cabinet

The light intensity that can come from various light sources in the outdoor atmosphere can decrease the precision of image processing algorithms. In this case, the cabinet is set up to minimize the difference in exterior light intensity. [21] The equipment with a reduced light transmission index protect the quality control process. With the aid of Keyence, the light control equipment, image processing unit, cabin calculations, and camera positions are determined.

The object tracking device isn't in use, and it's also necessary to make sure the product is still in the same place. In this kind of image processing program, it's crucial. [23] [24] Parts must be traceable from the moment they are made in the IMM. Special fixtures are created for the portability and traceability of these pieces in this sense. This fixtures are used to create production flow and storage. Station-based traceability is often controlled by installing RFID devices under these fixtures. Product traceability is maintained in this manner. When any quality problem occurs, it can be controlled by direct intervention. These fixtures are assembled on the conveyor and continuously turn and come back from the beginning of the line to the end.

Figure 15: RFID Fixture

When these RFID fixtures enter the cabinet, they are locked by the automatic fixation guides on the conveyor. The fixture position is controlled using the switch and sensors on the automatic fixing guides. If the position is not correct, feedback data to the conveyor motor's drive provides accurate positioning within the cabinet. Unless there is approval from the automatic fixing guidance system, the camera control system will not work.

3.2.2 Image Processing for Quality Control System

OpenCV is a free open-source library that includes a plugin for image processing. C/C++ are used to build this program. Effective programs are the product of code architecture and working systems. Many high-level programming languages, including Python for images, are supported by the library. Scikit-image is a Python programming language framework with a necessary and practical solution logic.

Standard applications of OpenCV's image processing module are for two dimensional images. [25]

OpenCV is free to use commercially and is open source. OpenCV is a highly structured library for real-time applications that works with the Python programming language. The image window would not match on the screen if the image volume is much bigger than the region to be observed. The entire image is viewed using OpenCV's cv2.WINDOW NORMAL approach, and the image can be resized. Figure 4.8 shows an image that is set to reflect on the piece's possible short shot area.

Figure 16: Potential Short Shot Quality Area

A grayscale image is used to define edges in this project. Grayscale images can be used to identify and recognize traces of contrast or shading in an image, as well as to analyze the image's specification. On RGB images with three channels: Red, Green, and Blue, grayscale image solutions have two-dimensional channels. A pixel in a grayscale image has less data than a pixel in a color image. In terms of processing, it is very convenient. CvtColor is used to transform an RGB image to a grayscale image in this project ().

Figure 17: Gray scaling the color image

The template Matching approach is typically used to spot a template within a source image. When it comes to separating objects, the approach isn't very successful. The Sliding window algorithm searches the template that was found on the source image using the Template Matching method. The template is rotated on all pixels in (1,1) coordinates on the source, a similarity ratio is calculated using the similarity form the choose, and if the template and the currently rotated template are identical, it returns those pixels. In this analysis, template matching is used. For this, OpenCV provides cv2.matchTemplate ().

The method basically focuses on a certain region in the large image and shifts the template to that location. The amount of overlap between each pixel in the image and the template is computed. Consequently, there is a similarity between the template and the image.

Figure 18: Template, Normal Production.

The picture on the left is a template, the picture on the right is suitable for quality. The template matching similarity score is %95,7. Generally, production is taken under these conditions.

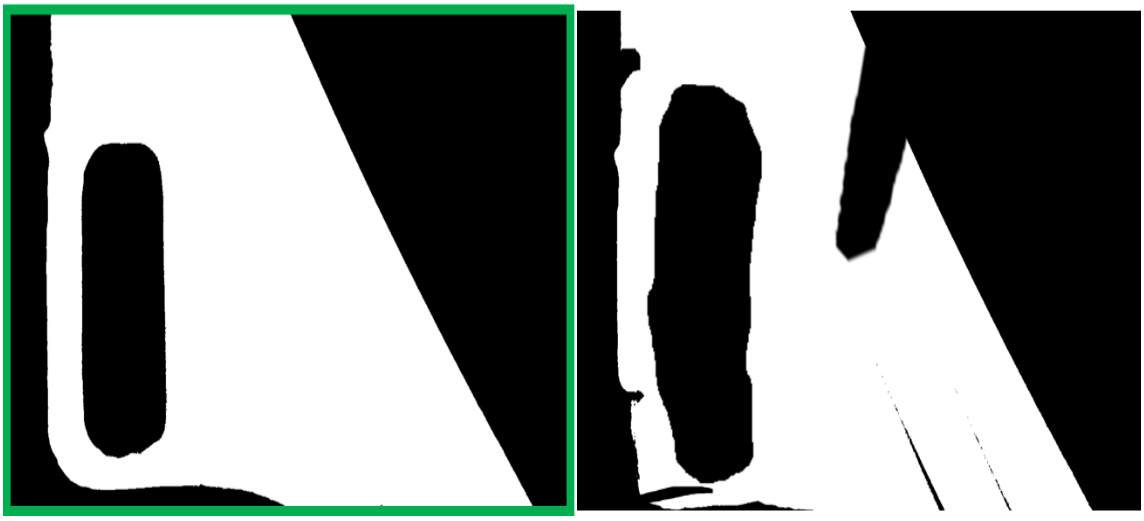

Figure 19: Template, Short Shot Quality Failure

The picture on the left is a template, the image on the right is short shot quality failure. The template matching similarity score is %75,3. In normal conditions, it is not acceptable to have an incomplete injection error on the part surface. In this context, the threshold is determined as 90% due to the meetings with the quality department.

3.2.3 Model Result

An imbalanced dataset is a form of data distribution in which one or more classes have the most instances in comparison to other classes. In this case, the confusion matrix makes life easier. The confusion matrix efficiency metric for an imbalanced dataset is efficient and straightforward. Also, provide detailed details about the goal.

Figure 20: Confusion Matrix

Determining the most suitable performance criterion to use for classifying class imbalanced data sets is a significant challenge. This kind of imbalance dataset can benefit from the Matthews Correlation Coefficient. This method produces a comparable outcome. [26] The researchers are looking for a variety of statistical findings to use in assessing binary classification accuracy. One of the most often used outcome metrics in the binary classification methodology is the F1 score and uncertainty matrix. When evaluating imbalance datasets, these predictive outcome indicators can be risky. [27]

An alternative solution strategy is to use an alternative Matthews correlation coefficient that is unaffected by imbalanced datasets. This method is described in detail below which form can also be help.

Depending on if M is used, the MCC looks like this:

$$M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix} \tag{7}$$

$$MCC = \frac{TPxTN - FPxFN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \tag{8}$$

(worst value: –1; best value: +1)

If MCC can correctly estimate the vast majority of positive and negative data samples in an imbalance dataset, it can receive a high score. It produces results in the range [-1, + 1]. MCC +1 denotes a perfect outcome. It will generate a total of 0 random responses. The inverse predictor is shown by -1. According to this approach, MCC value is calculated as +0.54. [27]

3.3 Summary

The IMM machine data infrastructure is completed with the OPC-UA communication protocol and MES infrastructure. The camera control system is activated taking into account the light intensity, exposure settings and part position. The image processing model success metric is evaluated. In studies with weekly controls, it is seen that the Matthews correlation coefficient is higher than +0.5 and the model recall value is higher than 80%. With the establishment of the camera control station, short shot quality defect decreases from 3.7% to 2.45%.

CHAPTER 4

DIGITAL TRANSFORMATION

In this section, the concept of digital transformation is introduced in Industry 4.0 by Kotarba, Schallmo, Demartini, Weiblen, and He. The studies on the visualization of IMM parameters transferred to MES within the logging and digital twin scope are explained. Studies conducted within the scope of modeling short shot quality error with ML are mentioned (Fabian, Krawczyk). In these studies, the CRISP-DM approach is followed for the industry mentioned by Huber. It is stated how the developed model is applied in the field and how to create a strategy for deployment.

4.1 Virtual Presentation of the Process

The concept of "digital transformation" has recently become a hot topic in the manufacturing industry with the advent of Industry 4.0. In this context, digitalization means the adaptation of new business models, inviting technological progress, and calling for innovations for the manufacturing environment. Digitalization also enables manufacturing firms to rapidly adapt to changing circumstances [28] and transform manufacturing companies' processes and business models by allowing new product and service offerings while requiring new business partnerships [29]. This perspective challenged the traditional approaches to efficiency improvement. It forced manufacturing companies to build external collaborations with technology developer companies and startups to adopt new digital technologies that they are not yet competent.

Moreover, with a rapidly increasing need for sustainability performance, production process efficiency and low raw material consumption become the primary concerns for manufacturing companies. In this context, digital technologies' adoption gains importance since they deliver benefits for improving sustainability performance by optimizing the material and energy consumption and hence providing resource efficiency [30]. In particular, digital twin applications also enable sustainable, intelligent manufacturing. [31]

On the other hand, contemporary approaches such as open innovation, design thinking, and agile project or partner management also act as drivers and enablers of successful digital technology adoption by enabling companies to move beyond their digital capabilities and overcome rigid organizational limitations practices not-invented-here syndrome.

Though large manufacturers have advanced resource bases, they still need to expand their digital capabilities with entrepreneurial talent. On the other hand, startups can offer digital capabilities and assume a consultative role in guiding and mobilizing large companies [32]. They also have an essential edge over big firms in terms of agility. Hence, both sides try new ways to collaborate more effectively through agile practices [33].

4.1.1 Digital Twin

Developing a digital twin by simulating injection machinery molding (IMM) physics requires experience in the plastic injection process and complex simulation expertise. The team needs to understand the complex behavior of a production process to improve quality and estimate the material volume needed for process modeling. The project team has consulted with experienced operators and production engineers and analyzed how operators control the process and how operators overcome the quality problems. In this way, the team could maintain user involvement and sustain the continuous information flow from the real environment.

4.1.2 Visualization of Process Parameters

IMM machine manufacturer company which is Chen Hsong interviewed within the scope of data acquisition of process parameters. In this context, the manufacturer makes the process parameters accessible for data acquisition. Currently, these process parameters can be accessed from operator control SCADA PCs. However, the aim here is to ensure that the experienced operator management team can access these data in real-time and to prevent errors due to incorrect operator parameters. In each product cycle, these parameters are refreshed in real-time and logging in the ODS DB.

The data that could reach in the plastic injection process is evaluated by examining the sources. While these data are sufficient for us in the industry approach, the current parameter approach limited us in the Industry 4.0 concept from which the technology came. [34], [35]

To divide the IMM process into three main stages by examining the sources. These are the Injection Unit, Molding Unit, and Clampint Unit. We determined the parameters related to these sections and evaluated the short shot quality error aspects. [36]

The approach methods and their results to solve short shot quality error with different ways are evaluated. And finally, met with Chen Hsong company in this context and determined the following parameter blocks with the machine manufacturer.

Process Parameters:	***Value:***	***Unit:***	***Section:***
Decompress time	0.19	sec	INJ
Ejector backward time	0.84	sec	INJ
Ejector forward time	1.03	sec	INJ
Cycle time	88.2	sec	INJ
Injection end position	32.3	mm	INJ
Injection position 1	1	mm	INJ
Injection position 2	90	mm	INJ
Injection position 3	140	mm	INJ
Injection position 4	190	mm	INJ
Injection position 5	250	mm	INJ
Injection position 6	315	mm	INJ
Injection position 7	380	mm	INJ
Injection position 8	420	mm	INJ
Injection position 9	470	mm	INJ
Injection position 10	480	mm	INJ
Injection pressure 1	0	bar	INJ
Injection pressure 2	60	bar	INJ
Injection pressure 3	95	bar	INJ
Injection pressure 4	110	bar	INJ
Injection pressure 5	140	bar	INJ
Injection pressure 6	150	bar	INJ
Injection pressure 7	150	bar	INJ
Injection pressure 8	150	bar	INJ
Injection pressure 9	150	bar	INJ
Injection pressure 10	150	bar	INJ
Injection Speed	58	mm/s	INJ
Injection switchover position	109	mm	INJ
Injection time	10.7	sec	INJ
Injection velocity 1	0	mm/s	INJ
Injection velocity 2	7	mm/s	INJ
Injection velocity 3	16	mm/s	INJ
Injection velocity 4	23	mm/s	INJ
Injection velocity 5	29	mm/s	INJ
Injection velocity 6	38	mm/s	INJ
Injection velocity 7	47	mm/s	INJ
Injection velocity 8	52	mm/s	INJ
Injection velocity 9	55	mm/s	INJ
Injection velocity 10	58	mm/s	INJ
Max injction speed	63	mm/s	INJ
Max open position	1753	mm	INJ
Plastizing volume	517	gr	INJ
Shot Volume	503	gr	INJ

Figure 21: Injection Section Parameter

Process Parameters:	***Value:***	***Unit:***	***Section:***
Cooling temperature	34.9	°C	MOLD
Cooling time	47.3	sec	MOLD
HotRunner Zone 1 temperature	249	°C	MOLD
HotRunner Zone 2 temperature	244	°C	MOLD
HotRunner Zone 3 temperature	251	°C	MOLD
HotRunner Zone 4 temperature	246	°C	MOLD
HotRunner Zone 5 temperature	250	°C	MOLD
HotRunner Zone 6 temperature	245	°C	MOLD
HotRunner Zone 7 temperature	241	°C	MOLD
HotRunner Zone 8 temperature	244	°C	MOLD
HotRunner Zone 9 temperature	247	°C	MOLD
HotRunner Zone 10 temperature	242	°C	MOLD
HotRunner Zone 11 temperature	243	°C	MOLD
HotRunner Zone 12 temperature	239	°C	MOLD
HotRunner Zone 13 temperature	241	°C	MOLD
HotRunner Zone 14 temperature	240	°C	MOLD
HotRunner Zone 15 temperature	244	°C	MOLD
HotRunner Zone 16 temperature	247	°C	MOLD
HotRunner Zone 17 temperature	251	°C	MOLD
HotRunner Zone 18 temperature	250	°C	MOLD
HotRunner Zone 19 temperature	248	°C	MOLD
HotRunner Zone 20 temperature	246	°C	MOLD
HotRunner Zone 21 temperature	248	°C	MOLD
HotRunner Zone 22 temperature	247	°C	MOLD
HotRunner Zone 23 temperature	238	°C	MOLD
HotRunner Zone 24 temperature	246	°C	MOLD
Nozzle temperature	251	°C	MOLD
Tub_Temperature	58	°C	MOLD
Zone 1 temperature	223	°C	MOLD
Zone 2 temperature	229	°C	MOLD
Zone 3 temperature	224	°C	MOLD
Zone 4 temperature	231	°C	MOLD
Zone 5 temperature	224	°C	MOLD
Zone 6 temperature	228	°C	MOLD
Zone 7 temperature	229	°C	MOLD
Zone 8 temperature	230	°C	MOLD
Zone 9 temperature	228	°C	MOLD

Figure 22: Molding Section Parameter

Process Parameters:	***Value:***	***Unit:***	***Section:***
Carriage backward time	6.6	sec	CLAMP
Carriage forward time	1.5	sec	CLAMP
Clamp close time	3.6	sec	CLAMP
Clamp force	819	kN	CLAMP
Clamp open time	5.36	sec	CLAMP
Holding Position	99.7	mm	CLAMP
Holding time	6.21	sec	CLAMP
Metering time	29.1	sec	CLAMP

Figure 23: Clamping Section Parameter

An operator dashboard is prepared for the traceability of the IMM machine with the collected data. This dashboard can be followed from mobile, tablet, and PCs within the network. In the first place, the infrastructure is provided with the dashboard so that the production team and the management team can monitor the production.

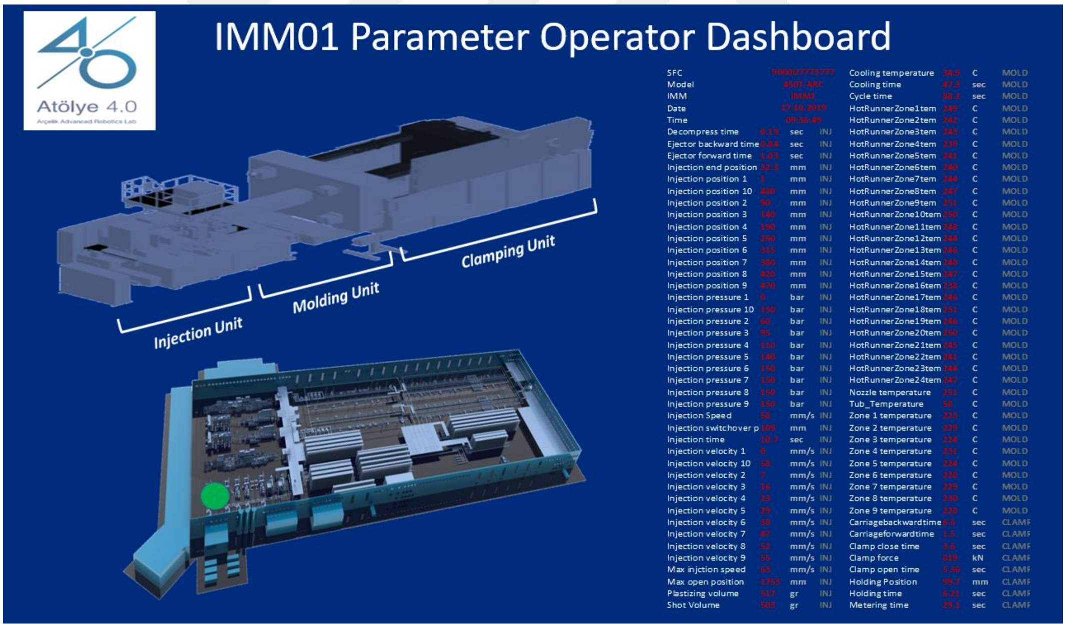

Figure 24: Operator Dashboard

Primarily, it has increased the learning process for the operators with little experience. They have improved their ability to control the production line. With this dashboard, the short shot quality failure has decreased by 6%.

4.2 Predictive Short Shot Quality Failure via Data Analytic

In the context of machine learning, Scikit-learn is one of the most commonly used libraries. This library has a wide range of applications, including help vector machines, logistic regression, gradient boosting, decision trees, and random forests. There are many explanations for the popularity of the scikit-learn package. The first is that it contains the majority of the fundamental processes. Second, scikit-learn makes it possible to run data analytics programs from beginning to end. Separate modules to fill in missing values in the data, select attributes, conduct cross validation, and analyze results eliminate the need for another package.

The positive thing about the scikit-learn program is that it has a straightforward API that eliminates the need to memorize different syntaxes with different methods. This allows beginners to data analytics to get their hands dirty easily. Using similar functions in different stages, such as clustering, regression, decision trees, filling missed values, and scaling data, makes the job much easier. The API is not only plain, but it also has excellent documentation. While, in terms of efficiency, libraries like lightgbm and tensorflow are ahead. [37]

4.3 Classification in Supervised Machine Learning for Quality

The CROSS Cross-Industry Standard Processing Process for Data Analysis is known as the CRISP-DM model (CRISP-DM). This is a robust data mining method model that has been reviewed. Before the business method becomes systematic, CRISP-DM simply illustrates the procedures and required measures from understanding the data to applying the solution. It is the most widely adopted theoretical model, as well as an open standard method model that expresses traditional data analysis approaches. [38]

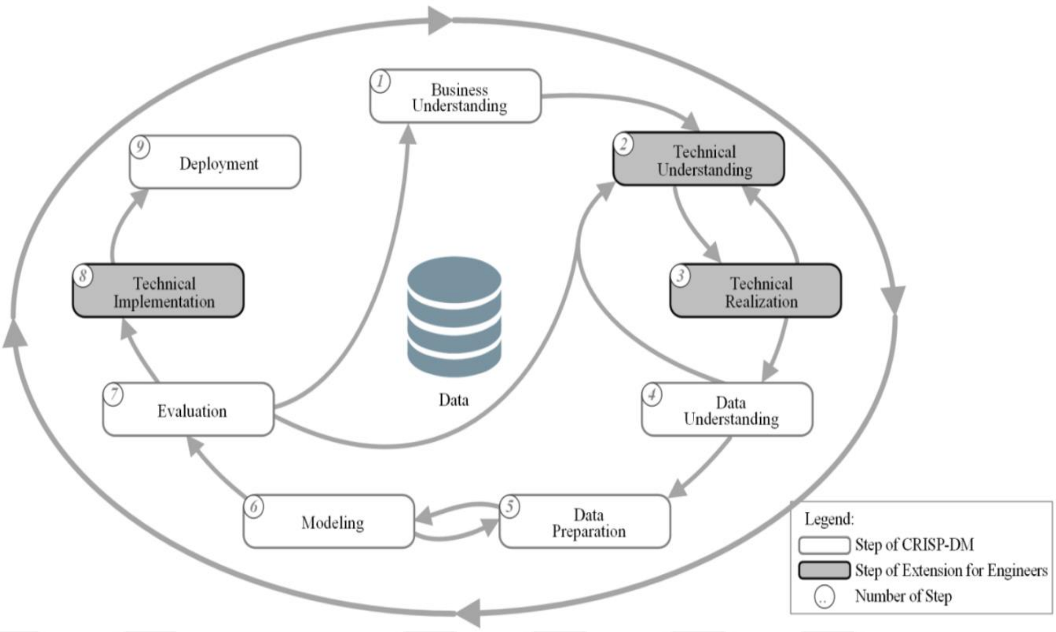

Figure 25: The DMME Process (Source: Huber[38])

4.3.1 Business Understanding

Our aim here is to introduce a model that will detect short shot quality defect errors during production and warn the operator early. The biggest problems encountered in production are the inexperienced operators not reading the production and having serial errors. Previously, this error could go up to the customer and cause serious product accidents. With the camera installed at the end of the line, the image processing system warns the operator when it encounters an error to prevent it from going to the customer.

The outcome of models created with classification algorithms, assessing the model's performance solely on the basis of accuracy rate can lead to a significant error in an imbalance collection of data. Predictive analytics is complicated by imbalance classification issues. The majority of classification algorithms are built on the assumption that each class has an equal number of samples. Because the rate of observation in each class is disproportionately high in imbalanced groups. [39]

As a result, models with poor prediction efficiency, especially for the minority class, are produced. This is a significant issue for this initiative. Since the minority class is usually more relevant. As a result, the minority class is more vulnerable to classified mistakes than the dominant class. Many machine learning algorithms can not provide accurate results when dealing with imbalanced datasets because they ignore the unequal distribution of classrooms. In other words, by categorizing the weighted class and attempting to reduce the overall error rate, the received classifiers can skip the minority class. [40]

IMM detects variability in process parameters and sets up an early warning system for the operator. Here, the aim is to increase the recall value above 80% in the predictive quality model and to calculate the MCC value above +0.50.

As a result, it is aimed to reduce the short shot quality error by 60% within the first 6 months after the model is commissioned.

4.3.2 Technical Understanding

It is aimed to log critical production data collected in IMM and label them with images in the camera control system. It is created to model the quality output in the approach of supervised machine learning algorithms with critical production data logged during production. It is aimed to model the quality output by establishing the classification technique.

The quality model created will be verified in production, and the results will be recorded. In this way, ML models' sustainability, which is one of the most critical issues in production, will be under control.

4.3.3 Technical Realization

As mentioned in 3.1.2 Technical Architecture in the previous section, critical process parameters are determined by first meeting with IMM manufacturer Chen Hsong. IMM consists of three main parts. These are injection, molding and clamping units, respectively. The machine has control units and PLC that control these main units and other auxiliary units. These PLCs are controlled by a master Beckhoff PLC. Master PLC is directly connected to the SCADA PC. The operator interface screen informs the operator on this SCADA PC, and process settings are made on this screen.

As stated in the previous section 3.2 Quality Control System, the quality control system is installed at the end of the line. In this quality control system, the potential short shot quality defect regions determined by the mold flow analyzer are photographed with Keyence VJ Series cameras. The captured images are processed with image processing and recorded.

Then, the processed data regarding the quality of the product is transferred to the SCADA PC with the OPC-UA infrastructure. Thus, the production data is labeled with the quality output. These logged data are integrated into the MES system and transferred to SAP-PCo, SAP-MII, and SAP-ME, respectively. Here, the SQL database with ODS-DB for IMM is saved. With a web service created, an infrastructure has been prepared for robotic process automation applications.

These logged data are accessed via web service, and the data is prepared for the machine learning approach in a CPU server located at a separate location. With data preprocessing, the data is made suitable for processing, and the results of ML algorithms are evaluated. And as a result, the model is commissioned on the SCADA PC, and it informs the operator on the warning screen prepared in the first place. A feedback system is then set up in the machine master PLC, which has been proven to be successful and commissioned.

4.3.4 Data Understanding

During the negotiations with Chen Hsong company, production process parameters related to making production adjustments have been determined until this time. The manufacturer reports that it can share the setting parameters in the conventional production infrastructure. These production parameters have been collected in 3 stages, as stated in the previous sections. These are the process parameters taken from the injection, molding, and clamping units.

In the Injection unit, the operator can adjust the Injection Position (Step: 1-10), Injection Pressure (Step: 1-10), and Injection Speed (1-10) values as the process set. Process times and Plasticizing Volume value are subtracted from mold flow analysis made according to the part to be produced. The ideal setting is tried to be found depending on the raw material, mold, and environmental conditions in production.

In the Molding unit, the operator can set the values of Hotrunner Temperatures (Section: 1-24), Barrel Temperatures (Section: 1-9) as process settings. The product temperature, cooling temperature, and cooling time change according to the behavior of the fluid in the mold. Hot runner and Barrel Temperature are extracted according to MFI technical specifications of the plastic raw material. According to the mold flow analysis and production domain expert information, the behavior of the product in the mold is one of the most critical features for short shot quality defects. Therefore, this molding unit is evaluated in detail.

It can adjust process time values as process setting in the clamping unit. Holding position and clamp force change according to the cooling behavior of the fluid in the mold. This part is critical for removing the substance from the mold without warping it.

4.3.5 Data Preparation

Numpy, pandas, sklearn, os, matplotlib, hyperopt, seaborn, imblearn, statsmodels, alert, math, xgboost, and lightgbm are among the Python modules used. To call the web server, the data is obtained in JSON format. After calling the web server, a script is written to parse the JSON results.

```
def raw_data_pre(start_date_2):

    now = datetime.now() - timedelta(hours=24)
    start_date = now.strftime("%Y-%m-%d")
    files = {start_date}

    for file in files:
        start_date = file
#     start_date = "2019-11-01"
# Limited with 1-day interval for query result!
        end_date = start_date

        start_time = "00:00:00.000"
        end_time = "23:59:59.999"

        #     data = json.loads(url.read().decode())

      with:urllib.request.urlopen("http://aculm20v.ar.arcelik:8000/XMII/Illuminator?Qu
      eryTemplate=Arcelik/ODS_IMM_Transfer/XAC_IMM_Data_Parameters&IN_StartDate="+star
      t_date+"%20"+start_time+"&IN_EndDate="+
      end_date+"%20"+end_time+"&j_username=MES_SYS_IMMDataParam&j_password=DataParam20
      19&Content-Type=text/json") as url:

        data = json.loads(url.read().decode())
        data = data["Rowsets"]
        df_raw = json_normalize(data['Rowset'], 'Row')
        df_dict = df_raw['Measures'].apply(lambda r: ast.literal_eval(r))
        df_values = pd.DataFrame()


        for i in range(len(df_dict)):
            temp = pd.DataFrame (df_dict[i]). set_index ("MeasureName").T
            df_values = pd.concat([df_values, temp], sort=True)

        df_values.index = range(len(df_values))
        df_raw = df_raw[["Datetime", "Item", "Operation", "Resource", "SFC"]]
        df_final = pd.concat([df_raw, df_values], axis=1, sort=False)
        df_final.to_csv("data" + start_date + ".csv")
```

Figure 26: Part of the Parsing JSON Data Code

Web service is transformed daily with robotic process automation with a data frame, and SQL is saved. The code block has been converted into API, and it has been provided to run as a time-adjusted service in the windows operating system. Thus, the data frame is created.

One of most powerful way to begin any classification problem is to use Exploratory Data Analysis to analyze the data collection. EDA (Exploratory Data Analysis) is a means of evaluating databases in order to summarize main elements, and often employs visual tools. The aim here is to generate as much data-related information as possible. When the data is examined, it is seen that it is imbalanced. This leaves us with something like 30:1 ratio between the good quality and short shot quality failure classes.

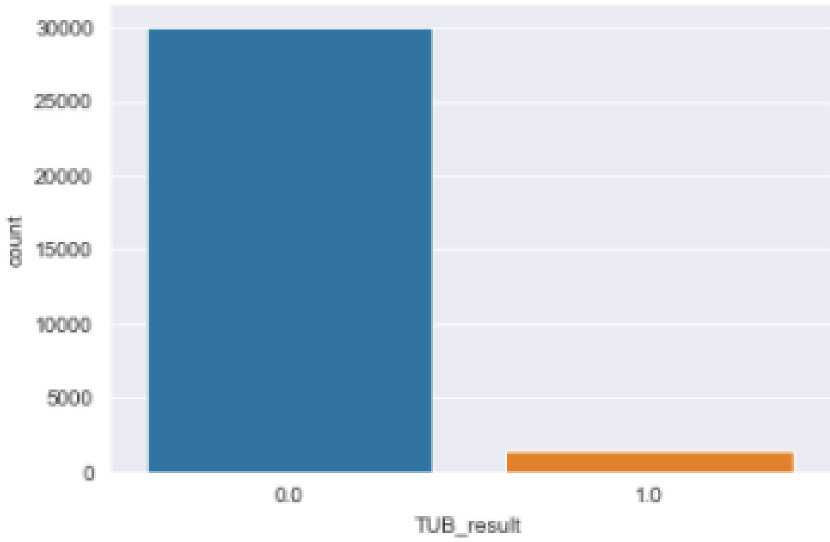

Figure 26: Imbalanced dataset

All features are arranged to be logged as float data type. When the dataframe is first examined, it is seen that the Injection Pressure (7-9), Hotrunner Temperature (22) and Decompress Time features are recorded as objects. These features have been converted to float. Then, by describing the data frame, count, mean, standard deviation, min, 25%, 50%, 75%, and max values are examined and data distribution behavior is reviewed.

When examining the data frame features, especially the temperature features collected from the mold unit are approached with the following approach. It is checked whether the incoming data is null or meaningless data one-time or serially. If there is one-time disambiguation, this null data is filled by taking the average of the value in the previous row and the value in the next row in the data frame. Because, temperature values cannot change dramatically between two cycles. Then the comparison of null and non-null is made according to the quality class. One hundred seventeen null rows have been detected. It is seen that the number of short shot quality defects in this 117-row dataset is 5.

In other features such as time, velocity, pressure, and displacement, the row has been deleted if the data is null.

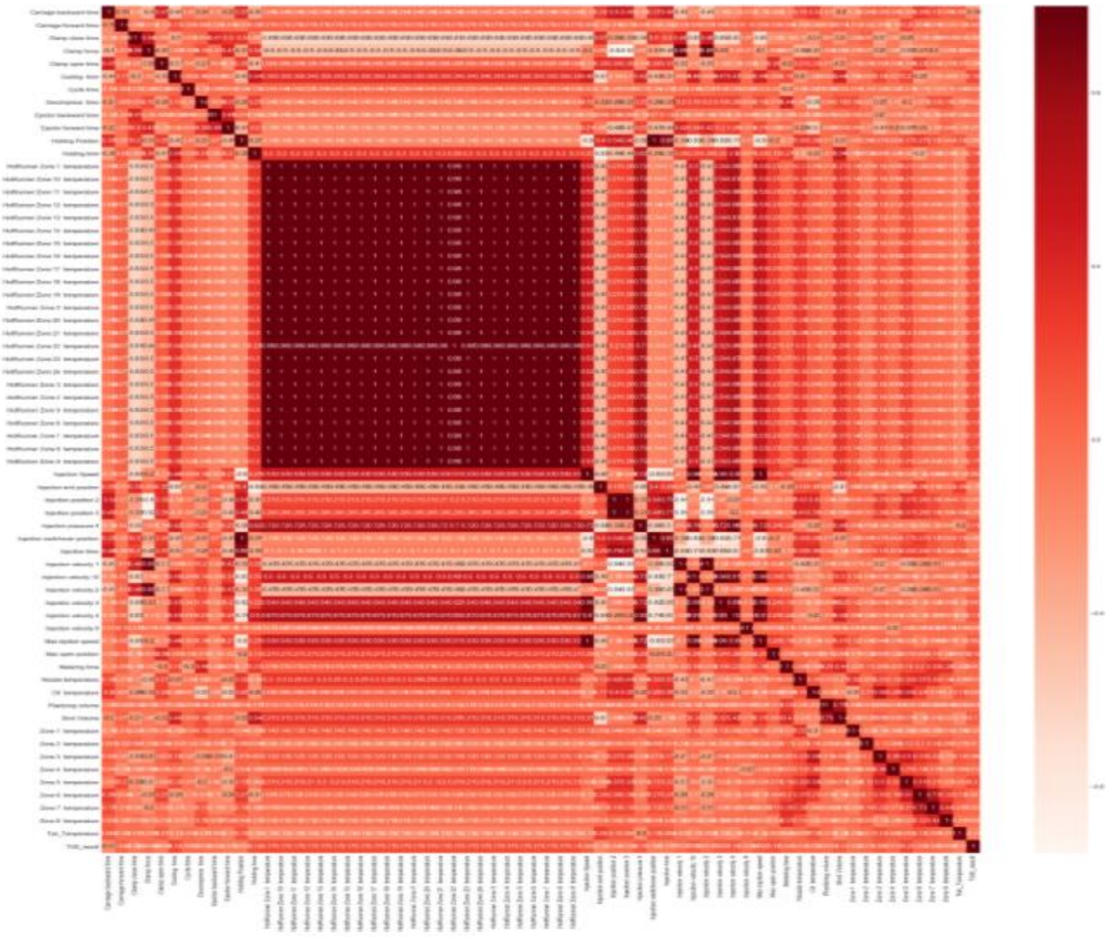

Figure 27: Pearson correlation coefficient

The Pearson correlation coefficient (PCC) is a mathematical instrument that determines how closely two variables are related. It produces a value between +1 and -1. A positive linear correlation is indicated by a value of 1, no linear correlation is indicated by a value of 0, and a negative linear correlation is indicated by a value of -1. In this case, association provides a framework for all evidence regarding the size and direction of correlations. [40], [41]

```
selected_time5 = ['Zone 1  temperature',
 'Zone 2  temperature',
 'Zone 3  temperature',
 'Zone 4  temperature',
 'Zone 5  temperature',
 'Zone 6  temperature',
 'Zone 7  temperature',
 'Zone 8  temperature']
```

```
df_zontemp_mean = df_time5.mean(axis=1)
df_zontemp_max = df_time5.max(axis=1)
df_zontemp_min = df_time5.min(axis=1)
df_zontemp_std = df_time5.std(axis=1)
df_zontemp_skew = df_time5.skew(axis=1)
df_zontemp_kurt = df_time5.kurt(axis=1)
```

```
selected_time1 = [ 'HotRunner Zone 1  temperature',
 'HotRunner Zone 10  temperature',
 'HotRunner Zone 11  temperature',
 'HotRunner Zone 12  temperature',
 'HotRunner Zone 13  temperature',
 'HotRunner Zone 14  temperature',
 'HotRunner Zone 15  temperature',
 'HotRunner Zone 16  temperature',
 'HotRunner Zone 17  temperature',
 'HotRunner Zone 18  temperature',
 'HotRunner Zone 19  temperature',
 'HotRunner Zone 2  temperature',
 'HotRunner Zone 20  temperature',
 'HotRunner Zone 21  temperature',
 'HotRunner Zone 22  temperature',
 'HotRunner Zone 23  temperature',
 'HotRunner Zone 24  temperature',
 'HotRunner Zone 3  temperature',
 'HotRunner Zone 4  temperature',
 'HotRunner Zone 5  temperature',
 'HotRunner Zone 6  temperature',
 'HotRunner Zone 7  temperature',
 'HotRunner Zone 8  temperature',
 'HotRunner Zone 9  temperature']
```

```
df_hotrunner_mean = df_time1.mean(axis=1)
df_hotrunner_max = df_time1.max(axis=1)
df_hotrunner_min = df_time1.min(axis=1)
df_hotrunner_std = df_time1.std(axis=1)
df_hotrunner_skew = df_time1.skew(axis=1)
df_hotrunner_kurt = df_time1.kurt(axis=1)
```

Figure 28: Feature Extraction from Hotrunner and Barrel Temperature

The feature's PCC value higher than + 0.9 directly removed. The features of Pearson correlation coefficient higher than + 0.5 are Hotrunner Temperatures and Barrel Temperatures. The approach of domain expert are extracted instead of these features. In this context, instead of these correlated features, the mean, max, min, std, skewness and kurtosis values of this data set are extracted.

```python
from __future__ import print_function

import numpy as np
from scipy.integrate import simps
from numpy import trapz

df['Dif_Vel'] = simps(df_vel, dx=4)
area = simps(df_vel.iloc[0], dx=4)
print("area =", area)
```

```
area = 378.0
```

```python
from __future__ import print_function

import numpy as np
from scipy.integrate import simps
from numpy import trapz

df['Dif_Pres'] = simps(df_pres, dx=4)
area = simps(df_pres.iloc[0], dx=4)
print("area =", area)
```

```
area = 1800.0
```

Figure 29: Feature Extraction from Injection Velocity and Injection Pressure

Other features with high PCC values are Injection Pressure and Injection Velocity values. Instead of these features, new features are extracted together with the domain expert. Explicit integration can be officially translated as a signed regional location on a plane tied to a graph of a given task between two points in a real line. Injection Velocity (1-10) and Injection Pressure (1-10) data here are considered as a function, and the underlying area has been added as a feature. In this context, this ten feature data set has been integrated.

```python
df["Cushion"] = df["Plastizing volume"] - df["Shot Volume"]
df["Diffential"] = df["Dif_Vel"] / df["Dif_Pres"] * df["Cushion"]
df["Shrinkage"] = df["Diffential"] *df["Tub_Temperature"]
df["Viscosity"] = df["Holding Position"] * df["Max injction speed"]
df["Thermal_Effect"] = (df["Zone_Mean"] - df["Tub_Temperature"]) /  df["Cooling  time"]
df["Thermal_Stress"] = df["Thermal_Effect"] * df["Max injction speed"]
df["Hydraulic_Force"] = (df["Zone_Min"] - df["Oil  temperature"]) / df["Cushion"]
```

Figure 30: Feature Extraction with Domain Expert

In the fields of image recognition, classification, and compression, PCA is a valuable mathematical method. It's a tool whose key goal is to retain a data set with high uncertainty and high-quality data while reducing the size of the data set. It allows for the reduction of the number of measurements and compression of data by identifying general features of multi-dimensional data.

PCA may be a very powerful tool for exposing the data's mandatory detail. PCA's basic logic is to mean a multidimensional data set with fewer variables by collecting the data's important features. [43]

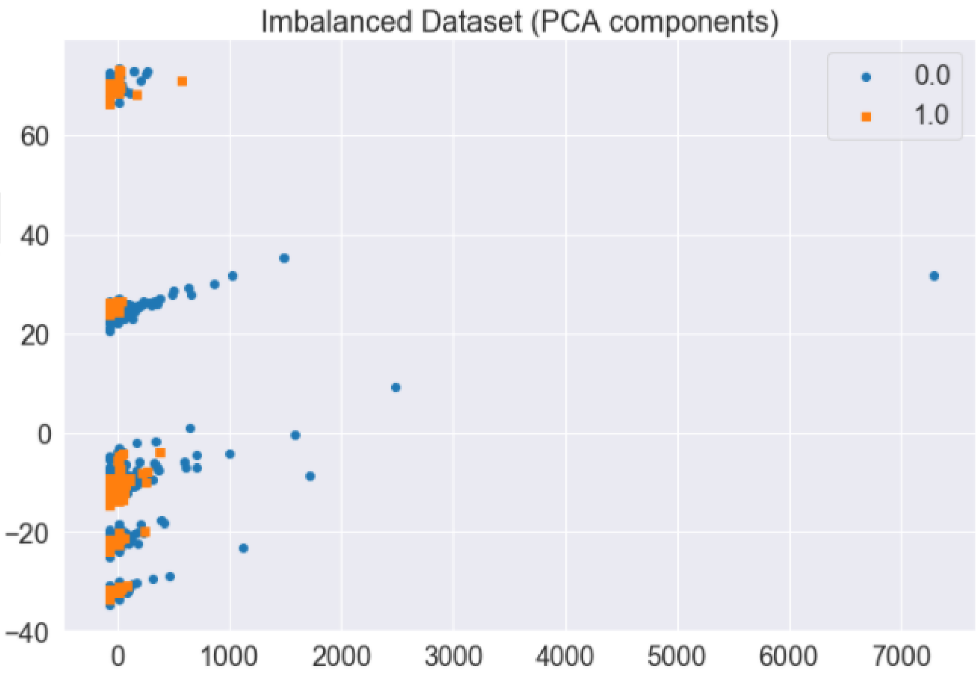

Figure 31: Imbalanced Dataset PCA Components

Through size reduction, certain characteristics will be lost, but the goal is that these lost traits will provide little knowledge about the population. This approach incorporates strongly correlated variables to create a smaller number of artificial variables known as "principal elements" that make up the majority of the data variance. The basic logic of PCA is to highlight multidimensional data with fewer variables by collecting the data's fundamental characteristics. [44] The PCA method must call from the sklearn library, in order to execute the PCA on the info package.

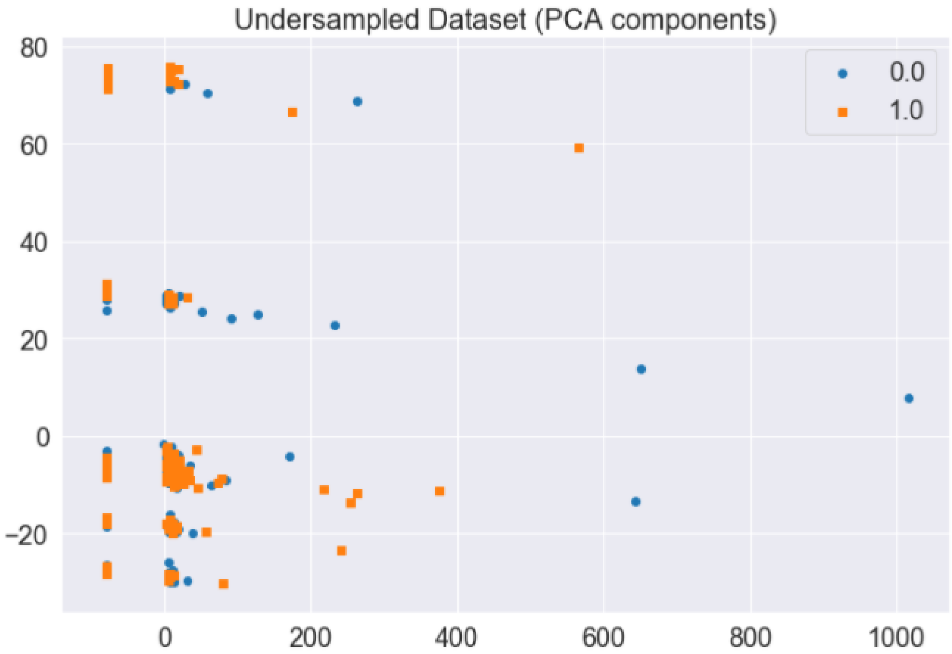

Figure 32: Undersampled Dataset PCA Components

We can stabilize imbalanced datasets by resampling. The first approach of this is by using different methods and groups with the same amount of data to sample data from the minority class (Class-1). For instance, Naive Random Resampling (Naive Random Over-Sampling) is used to achieve equilibrium by randomly selecting and copying existing data from the minority class. In addition, the equilibrium can be accomplished with the synthetic data generation using the interpolation process in SMOTE (Synthetic Minority Oversampling Technique) and ADASYN (Adaptive Synthetic Sampling Method) algorithms. [45], [46]

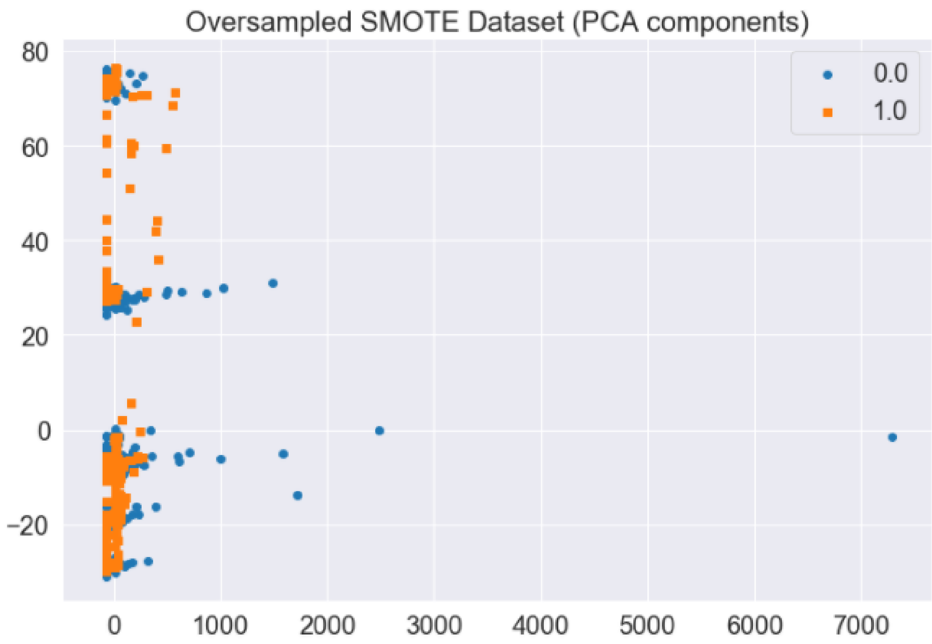

Figure 33: Oversampled SMOTE Dataset PCA Components

Data tested for a balanced 50:50 distribution in all methods in previous studies. The effectiveness of the method is increased by the identification of the best model sampling rate. Identify a potential sample ratio list and see the best performing mixture based on the average ROC AUC score after cross-validation. These values are chosen by random display and after a certain trial and error can be changed by other values. A resampled data for any pair of ratios together is then used to train the machine learning algorithm until the average ROC AUC score is issued after cross validation.

SMOTE oversampling rate:0.3,Random undersampling rate:0.7,Mean ROC AUC: 0.797
SMOTE oversampling rate:0.3,Random undersampling rate:0.6,Mean ROC AUC: 0.814
SMOTE oversampling rate:0.3,Random undersampling rate:0.5,Mean ROC AUC: 0.765
SMOTE oversampling rate:0.4,Random undersampling rate:0.7,Mean ROC AUC: 0.812
SMOTE oversampling rate:0.4,Random undersampling rate:0.6,Mean ROC AUC: 0.800
SMOTE oversampling rate:0.4,Random undersampling rate:0.5,Mean ROC AUC: 0.797
SMOTE oversampling rate:0.5,Random undersampling rate:0.7,Mean ROC AUC: 0.788
SMOTE oversampling rate:0.5,Random undersampling rate:0.6,Mean ROC AUC: 0.814
SMOTE oversampling rate:0.5,Random undersampling rate:0.5,Mean ROC AUC: 0.740

Figure 34: The mean ROC AUC for every possible pair of ratios combination

As seen in the output diagram, the best sample rate in this example was to have a sample value of 0.4, followed by an average of less than 0.7 for the best possible mean ROC AUC score at 81.2%. As a result, a 3150-row data frame is created.

The simulation paradigm is challenged by problems of unjust inequality, as most machine learning algorithms used for segmentation have been programmed to take the same sample number per class, since imbalanced groups have a disproportionate rate of detection throughout each class. This results in models with poor prediction performance, especially for the minority class. This situation is a major concern for the problem. The minority class is traditionally more significant. The concern is also more vulnerable than for the main class of classification errors for the minority class. Many machine learning algorithms could not be efficient when dealing with imbalance data sets, by missing the unequal classroom distribution. In other words, the classificators obtained will neglect the minority by classifying the weighted class in an attempt to reduce the overall error rate.

4.3.6 Modeling

Recursive Feature Elimination (RFE) is just kind of technique deletes the worst performing attributes step by step from the entire attribute collection to the better attribute subset. [47] This is a kind of tool for selecting reverse features. This method determines first all features based on an objective function and subtracts them from the feature set with the lowest score, in order to detect the feature subset that optimizes classification efficiency. This method goes on until the highest level of achievement is achieved. The functioning theory of the RFE algorithm is based on the training method with the classified machine learning algorithms, the calculation of attribute weights and the reduction of characteristics. [48]

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV
X = df_b1.drop('TUB_result', axis=1)
target = df_b1['TUB_result']

rfc = RandomForestClassifier()
rfecv = RFECV(estimator=rfc, step=1, cv=StratifiedKFold(10), scoring='roc_auc')
rfecv.fit(X, target)
```

```
RFECV(cv=StratifiedKFold(n_splits=10, random_state=None, shuffle=False),
      estimator=RandomForestClassifier(bootstrap=True, ccp_alpha=0.0,
                                       class_weight=None, criterion='gini',
                                       max_depth=None, max_features='auto',
                                       max_leaf_nodes=None, max_samples=None,
                                       min_impurity_decrease=0.0,
                                       min_impurity_split=None,
                                       min_samples_leaf=1, min_samples_split=2,
                                       min_weight_fraction_leaf=0.0,
                                       n_estimators=100, n_jobs=None,
                                       oob_score=False, random_state=None,
                                       verbose=0, warm_start=False),
      min_features_to_select=1, n_jobs=None, scoring='roc_auc', step=1,
      verbose=0)
```

```
print('Optimal number of features: {}'.format(rfecv.n_features_))
```

```
Optimal number of features: 11
```

Figure 35: Recursive Feature Elimination with Cross-Validation via Random Forest

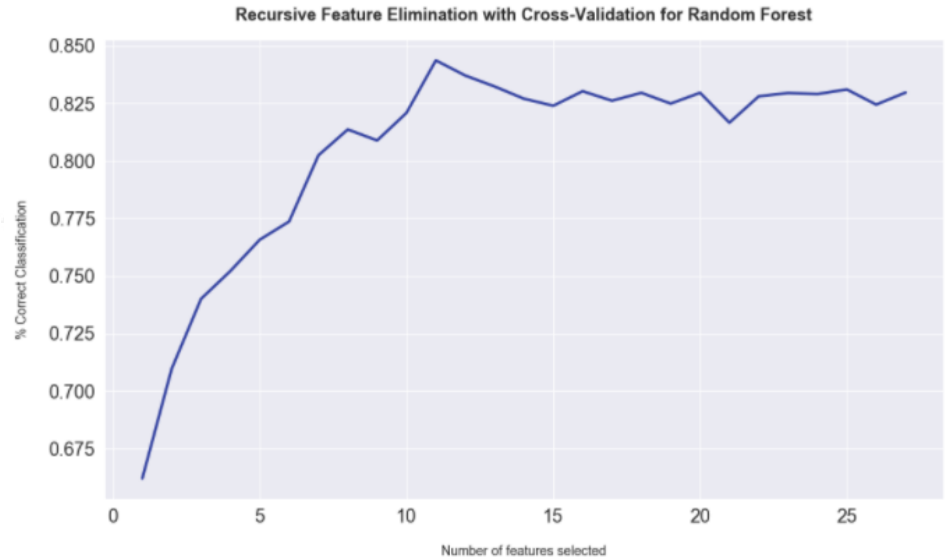

Figure 36: Number of Features Selected

In this example, we can see that 11 features were selected, though there doesn’t appear to be much improvement in the model's AUC score after around 7 features. The feature selection is a critical role in clarifying the outcome of the model performance metric. Being able to set up the model simply with fewer features will positively affect the test results.

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV
X = df_b1.drop('TUB_result', axis=1)
target = df_b1['TUB_result']

rfc = DecisionTreeClassifier()
rfecv = RFECV(estimator=clf, step=1, cv=StratifiedKFold(10), scoring='roc_auc')
rfecv.fit(X, target)
```

```
RFECV(cv=StratifiedKFold(n_splits=10, random_state=None, shuffle=False),
      estimator=DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None,
                                       criterion='gini', max_depth=None,
                                       max_features=None, max_leaf_nodes=None,
                                       min_impurity_decrease=0.0,
                                       min_impurity_split=None,
                                       min_samples_leaf=1, min_samples_split=2,
                                       min_weight_fraction_leaf=0.0,
                                       presort='deprecated', random_state=None,
                                       splitter='best'),
      min_features_to_select=1, n_jobs=None, scoring='roc_auc', step=1,
      verbose=0)
```

```
print('Optimal number of features: {}'.format(rfecv.n_features_))
```

```
Optimal number of features: 18
```

Figure 37: Recursive Feature Elimination with Cross-Validation via Decision Tree

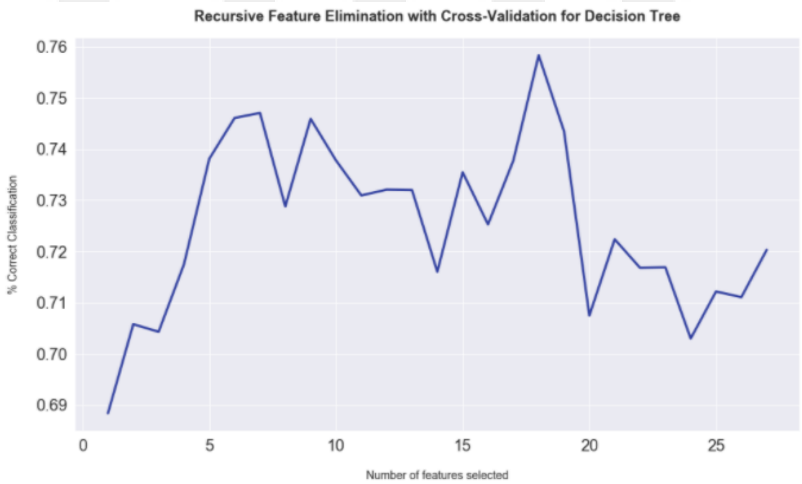

Figure 38: Number of Features Selected

The figure indicates a positive RFECV curve and when information is reached the curve leaps to best accuracy and decreases inaccuracies steadily as non-instructive characteristics are included in the model. The evaluation and comparison of these important functions.

```
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV
X = df_b1.drop('TUB_result', axis=1)
target = df_b1['TUB_result']

rfc = GradientBoostingClassifier()
rfecv = RFECV(estimator=sgb, step=1, cv=StratifiedKFold(10), scoring='roc_auc')
rfecv.fit(X, target)
```

```
RFECV(cv=StratifiedKFold(n_splits=10, random_state=None, shuffle=False),
      estimator=GradientBoostingClassifier(ccp_alpha=0.0,
                                           criterion='friedman_mse', init=None,
                                           learning_rate=0.1, loss='deviance',
                                           max_depth=3, max_features=None,
                                           max_leaf_nodes=None,
                                           min_impurity_decrease=0.0,
                                           min_impurity_split=None,
                                           min_samples_leaf=1,
                                           min_samples_split=2,
                                           min_weight_fraction_leaf=0.0,
                                           n_estimators=15,
                                           n_iter_no_change=None,
                                           presort='deprecated',
                                           random_state=None, subsample=1.0,
                                           tol=0.0001, validation_fraction=0.1,
                                           verbose=0, warm_start=False),
      min_features_to_select=1, n_jobs=None, scoring='roc_auc', step=1,
      verbose=0)
```

```
print('Optimal number of features: {}'.format(rfecv.n_features_))
```

```
Optimal number of features: 14
```

Figure 39: Recursive Feature Elimination with Cross-Validation by Gradient Boosting

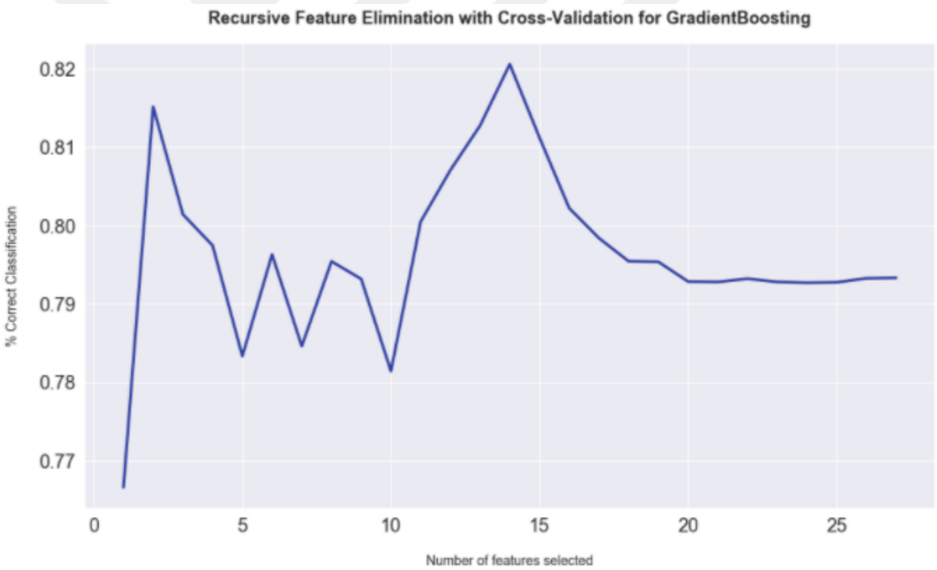

Figure 40: Number of Features Selected

With the feature elimination and dimension reduction approach methods, the 87 feature data frame is reduced to 13 features.

These features are Hotrunner_Temp_Max, Zone_Temp_Mean, Max_Injection_Pressure, Cooling_Time, Viscosity_M, Injection_Velocity_Integral, Max_Open_Position, Metering_Time, Nozzle_Temp, Hydraulic_Pump_Oil_Temp, Plastizing_Volume, and Shot_Volume.

The imbalanced data sets include training and testing. In order to enhance imbalance model performances steps, the data can also be skewed on whole data sets. [39]

Hyper-parameter tuning with grid and randomized methods is done in the following analysis. "RandomizedSearchCV / GridSearchCV."

```
# Create Decision Tree classifer object
clf = DecisionTreeClassifier()

# Train Decision Tree Classifer
clf = clf.fit(X_train, y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)


print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print ("AUC Score:", roc_auc_score(y_test, y_pred))
print ("Precision:", precision_score(y_test, y_pred))
print ("Recall:", recall_score(y_test, y_pred))
print ("F1 Score:", f1_score(y_test, y_pred))

feature_importances = pd.DataFrame(clf.feature_importances_,
                                   index = X_train.columns,
                                    columns=['importance']).sort_values('importance', ascending=Fal

print(feature_importances)
```

```
Accuracy: 0.7882736156351792
AUC Score: 0.7892872385830133
Precision: 0.7549668874172185
Recall: 0.8028169014084507
F1 Score: 0.7781569965870306
```

Figure 41: Decision Tree Model Performance

```
sgb = GradientBoostingClassifier(n_estimators=15)

sgb = sgb.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = sgb.predict(X_test)


print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print ("AUC Score:", roc_auc_score(y_test, y_pred))
print ("Precision:", precision_score(y_test, y_pred))
print ("Recall:", recall_score(y_test, y_pred))
print ("F1 Score:", f1_score(y_test, y_pred))

feature_importances = pd.DataFrame(sgb.feature_importances_,
                                   index = X_train.columns,
                                    columns=['importance']).sort_values('importance', ascending=Fal

print(feature_importances)
```

```
Accuracy: 0.8175895765472313
AUC Score: 0.8185232607767818
Precision: 0.7866666666666666
Recall: 0.8309859154929577
F1 Score: 0.8082191780821917
```

Figure 42: Gradient Boosting Model Performance

```
rfc = RandomForestClassifier(n_estimators=15)

rfc = rfc.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = rfc.predict(X_test)


print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print ("AUC Score:", roc_auc_score(y_test, y_pred))
print ("Precision:", precision_score(y_test, y_pred))
print ("Recall:", recall_score(y_test, y_pred))
print ("F1 Score:", f1_score(y_test, y_pred))

feature_importances = pd.DataFrame(rfc.feature_importances_,
                                   index = X_train.columns,
                                    columns=['importance']).sort_values('importance', ascending=Fal

print(feature_importances)
```

```
Accuracy: 0.8794788273615635
AUC Score: 0.8780623132735809
Precision: 0.8776978417266187
Recall: 0.8591549295774648
F1 Score: 0.8683274021352312
```

Figure 43: Random Forest Model Performance

```
xgb = XGBClassifier(objective='binary:logistic')
xgb.fit(X_train,y_train, eval_metric='auc')

#Predict the response for test dataset
y_pred2 = xgb.predict(X_test)


print("Accuracy:",metrics.accuracy_score(y_test, y_pred2))
print ("AUC Score:", roc_auc_score(y_test, y_pred2))
print ("Precision:", precision_score(y_test, y_pred2))
print ("Recall:", recall_score(y_test, y_pred2))
print ("F1 Score:", f1_score(y_test, y_pred2))
```

```
Accuracy: 0.8957654723127035
AUC Score: 0.8951771233461375
Precision: 0.8873239436619719
Recall: 0.8873239436619719
F1 Score: 0.8873239436619719
```

Figure 44: Xgboost Model Performance

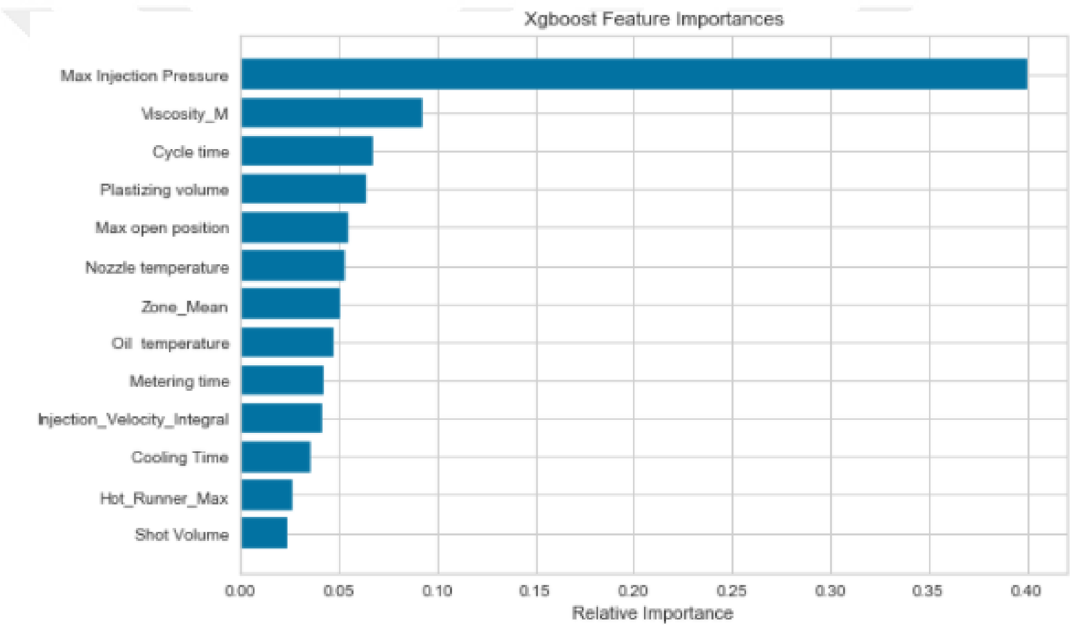

Figure 45: Xgboost Feature Importances

```
from xgboost import XGBClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import accuracy_score

param_grid = [{'min_child_weight': np.arange(0.1, 10.1, 0.1)}] #set of trial values for min_child_w
i=1
kf = StratifiedKFold(n_splits=5,random_state=1,shuffle=True)
for train_index,test_index in kf.split(X,y):
     print('\n{} of kfold {}'.format(i,kf.n_splits))
     xtr,xvl = X.loc[train_index],X.loc[test_index]
     ytr,yvl = y[train_index],y[test_index]
     model = GridSearchCV(XGBClassifier(), param_grid, cv=5, scoring= 'f1',iid=True)
     model.fit(xtr, ytr)
     print (model.best_params_)
     pred=model.predict(xvl)
     print('accuracy_score',accuracy_score(yvl,pred))
     i+=1
```

```
1 of kfold 5
{'min_child_weight': 4.0}
accuracy_score 0.8636363636363636

2 of kfold 5
{'min_child_weight': 6.1}
accuracy_score 0.88

3 of kfold 5
{'min_child_weight': 7.9}
accuracy_score 0.9371428571428572

4 of kfold 5
{'min_child_weight': 8.1}
accuracy_score 0.8971428571428571

5 of kfold 5
{'min_child_weight': 8.0}
accuracy_score 0.8971428571428571
```

Figure 46: Stratified K-Fold cross validation to obtain best model.

The test and error showed that min child weight contributed the most to improving the accuracy of the classifier by adding sklearn wrapper with, or without, the best values of other parameters.

4.3.7 Evaluation

After the model is created, the API is activated offline so that it does not give feedback to the operator. In this process, the results given to the products produced for a week are logged. The quality results given by the model as feedback and the actual production quality values are compared.

```
#Predict the response for test dataset
y_pred13 = xgb.predict(X_t)

print(confusion_matrix(y_t, y_pred13))
print("Accuracy:",metrics.accuracy_score(y_t, y_pred13))
print ("AUC Score:", roc_auc_score(y_t, y_pred13))
print ("Precision:", precision_score(y_t, y_pred13))
print ("Recall:", recall_score(y_t, y_pred13))
print ("F1 Score:", f1_score(y_t, y_pred13))
```

```
[[300 107]
 [  5  18]]
Accuracy: 0.7395348837209302
AUC Score: 0.7598547163764555
Precision: 0.144
Recall: 0.782608695652174
F1 Score: 0.24324324324324323
```

Figure 47: Test Result

According to these results, the AUC score is 92% and the Recall value is 88% in the quality predictive model development and initial test stages created from the xgboost classifier algorithm. When monitored in 1-week test production, the AUC score decreases from 92% to 75%. The recall value decreases from 88% to 78%. A reduction is expected during the testing phase. The results are above the initial targeted 70% recall and 75% AUC scores. Therefore, 1-month detailed production tracking is performed.

4.3.8 Technical Implementation

At this stage, it is aimed to put the model created on the SCDA PC together with the success metrics. The PyCharm program is used as the compiler. Libraries required in Python programming language are downloaded and prepared beforehand. The model is developed in the programming language of Python using joblib. This module is in the program sklearn. Built for Scikit-Learn models directly in serialization. It manages Numpy wide arrays better than other serialization modules for objects.

After this step, an API is created. In the program created, the web service prepared with RPA using the python programming language is activated on the backend side in each product cycle. Thus, raw data is prepared. Then this raw data is converted into the desired feature infrastructure with data frame preprocessing. The cycle is passed through the model and the result is logged. These data logged to SQL are designed to run as a service code created on the backend side.

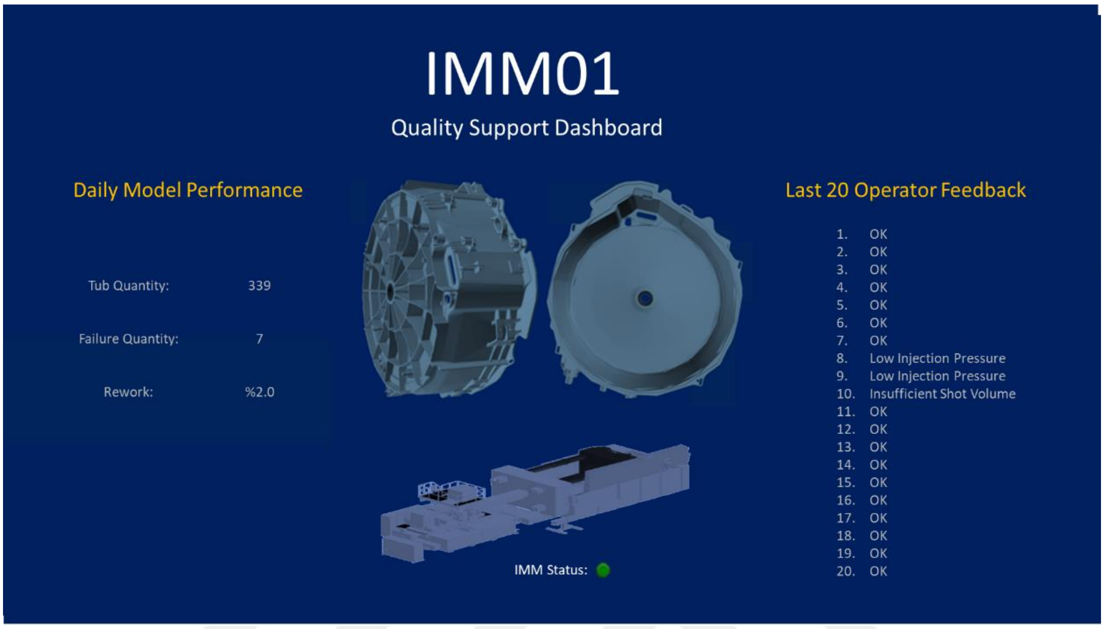

Figure 48: Quality Support Dashboard for Operator

This dashboard is prepared on node-red platform via node.js. On the left side of the dashboard interface prepared for the operator, under the heading of daily model performance, there are daily model production number, defective production number and rework rate. In the right-hand part, the logging of the feedback given during the last 20 productions is seen.

On the frontend side, this dashboard uses the node-red platform. IBM open-source node-red platform that has been developed by node.js. The root-node is an open source application; several modules have been created by users. The modules are referred to as nodes. The Node-Red program will create basic designs in the drag-and-drop logic of the ready nodes easily when developing the interface. On the opposite, changes can be rendered by node-red with JS codes if more technical design is needed.

And non-technicians may incorporate the desired minor improvements in a fast training gui.

4.3.9 Deployment

The system that gives a potential short shot quality defect warning during production and gives feedback to the operator for parameter changes is activated with the quality estimation model.

The model is monitored offline within the first 15 days after commissioning. Then, during the deployment phase, the system is monitored in production for about a month. During this period, a total of 9938 tub parts are produced. Short shot error is detected in 188 of these produced parts. The monthly rework rate is around 1.88%.

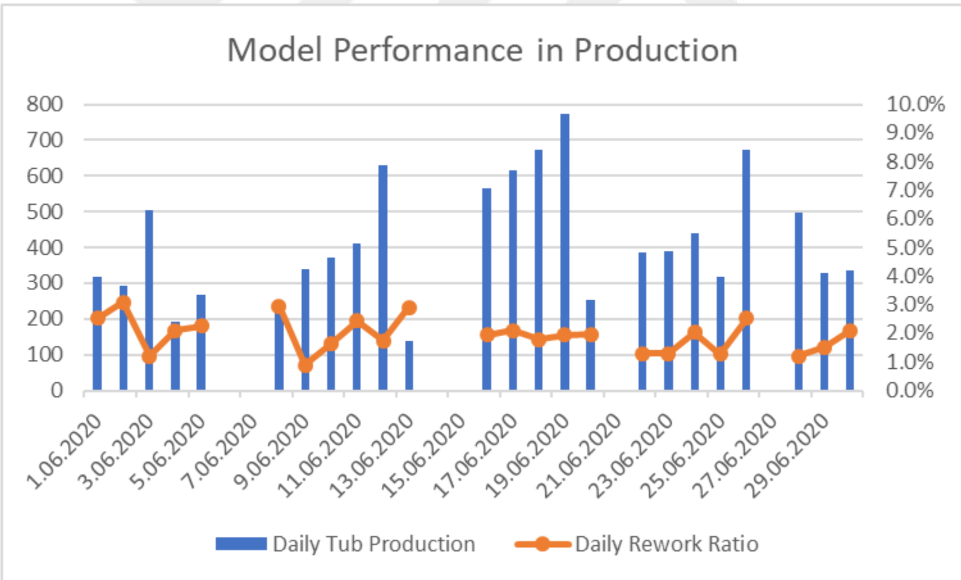

Figure 49: Model Performance in Production

The model is monitored offline within the first 15 days after commissioning. Then, during the deployment phase, the system is monitored in production for about a month. During this period, a total of 9938 tub parts are produced. Short shot error is

detected in 188 of these produced parts. The monthly average rework rate is around 1.84%. The average rework ratio is recorded as 1.88%, 1.94% and 1.73% respectively.

4.3.10 Deployment in Industry

The study has been done to prevent short shot error in a Tub model with a problem. During these studies, technical architecture is created by contacting the IMM machine manufacturer. Since IMM is in the MES system, it proceeds faster in terms of infrastructure. In addition, a camera control system investment is made at the end of the line. Cameras control the regions determined in the mold flow analysis. Photographs are processed with image processing and recorded in the system. These saved photos are matched with process parameters, and machine learning algorithms are run to prevent short shot error. The created model is activated in production, and feedback is given to the operator. This structure's establishment took eight months with the controls, although the system is ready for infrastructure. The system's sustainability has a negative impact on just one model due to such a long development time. Therefore, a deployment approach is created in production based on the work done.

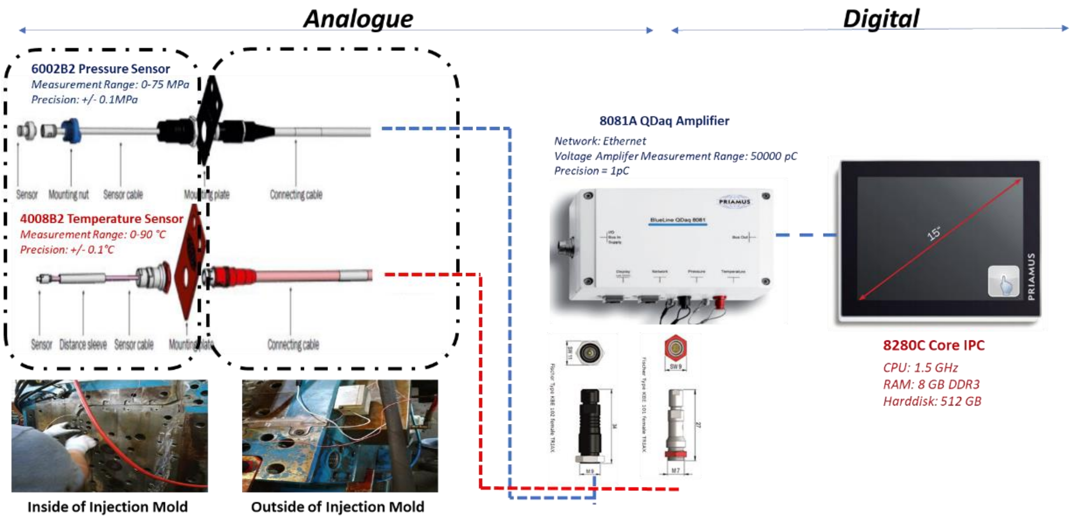

Figure 50: Techical Architecture of Inmold Sensor

As a result of the model created with the short shot, it is seen that the parameters collected from the mold unit are important in the feature importance scale. The data collected from the machine cannot give a time-series data about the behavior within the mold. Therefore, this information can be obtained with sensors to be placed in the mold. The information about what type of these sensors will be and where they will be placed is taken from mold flow analysis.

In this context, the flow behavior during injection can be modeled by installing pressure and temperature sensors in the determined parts of the mold. In this way, short shot quality defect can be prevented by installing sensors in the mold without any intervention to the IMM machine in the industry.

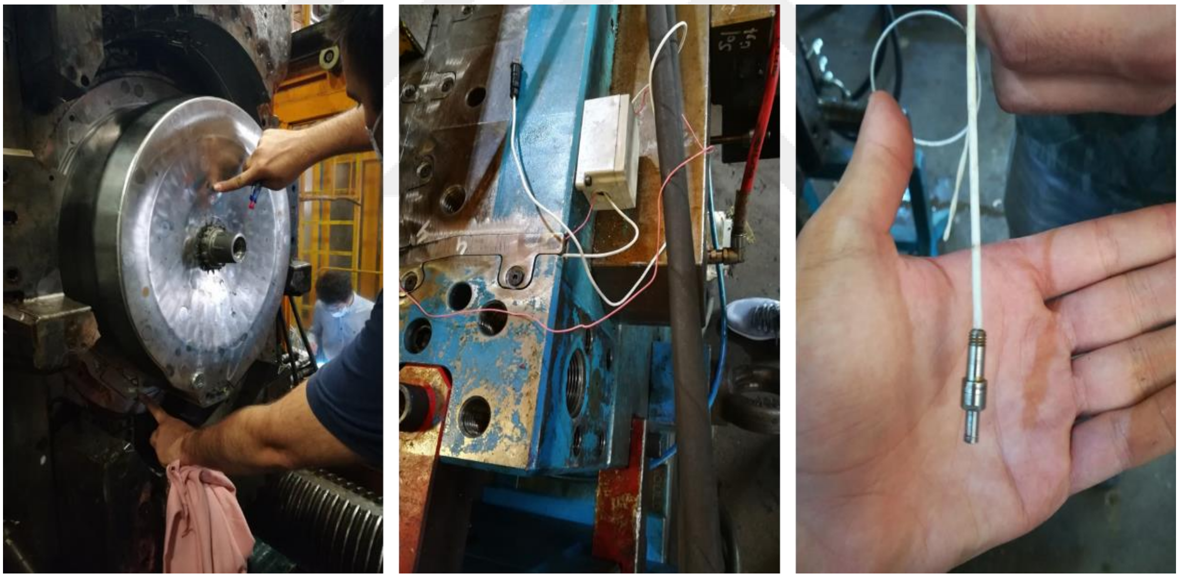

Figure 51: Sensor installation

Industrial research is carried out to disseminate this type of work in the industry. As a result of these industrial researches, Priamus company and its products are discussed. The cavity pressure sensors are based on the piezoelectric method of measurement. In order to ensure a given operating status and reset status, the systems are especially suitable for use during the injection molding process.

Thermocouple of type N are specially optimized for the temperature sensors. The thermocouple is selected since it does not display any permanent temperature shifts relative to other thermocouples with accurate advantages without corrosion. Like cavity strain, signal lines with secure ground connections and plugs are preferable.

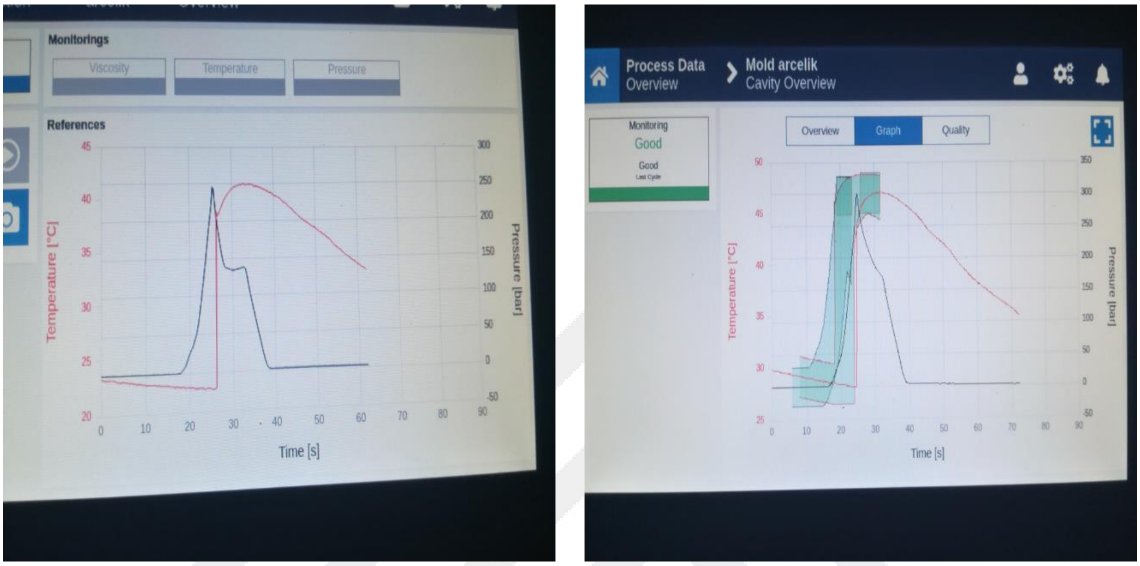

Figure 52: Inmold Pressure and Temperature Data Acquisition

With rising temperature, the flowability of most fluid plastics decreases. There are holes in the liquid, according to the hole principle, and molecules are continuously moving into spaces. This act allowed for flow, but it takes a lot of energy, requiring energy to move a molecule through orbit.

The more quickly activated electricity at high temperatures, the more easily fluid flows as temperature increases. On the other hand, the viscosity of a liquid increases as friction increases and the amount of voids in the liquid decreases under pressures. The flow of molecules is also hard to do.

The data in the mold starts to be collected. The feedback short shot quality system is activated by associating the collected pressure and temperature data with IMM critical process parameters. Since the system trials are in the first stage, the feedback system is set up as follows in the first place. Pressure and temperature safe ranges in critical regions are determined as a result of Moldflow analysis. Then the pressure data collected from the mold during production are IMM pressure set values (Step 1:10). Temperature data are activated by matching HotRunner Temp (1-24), Barrel Temp (1-9), and Valve Gate Opening Time.

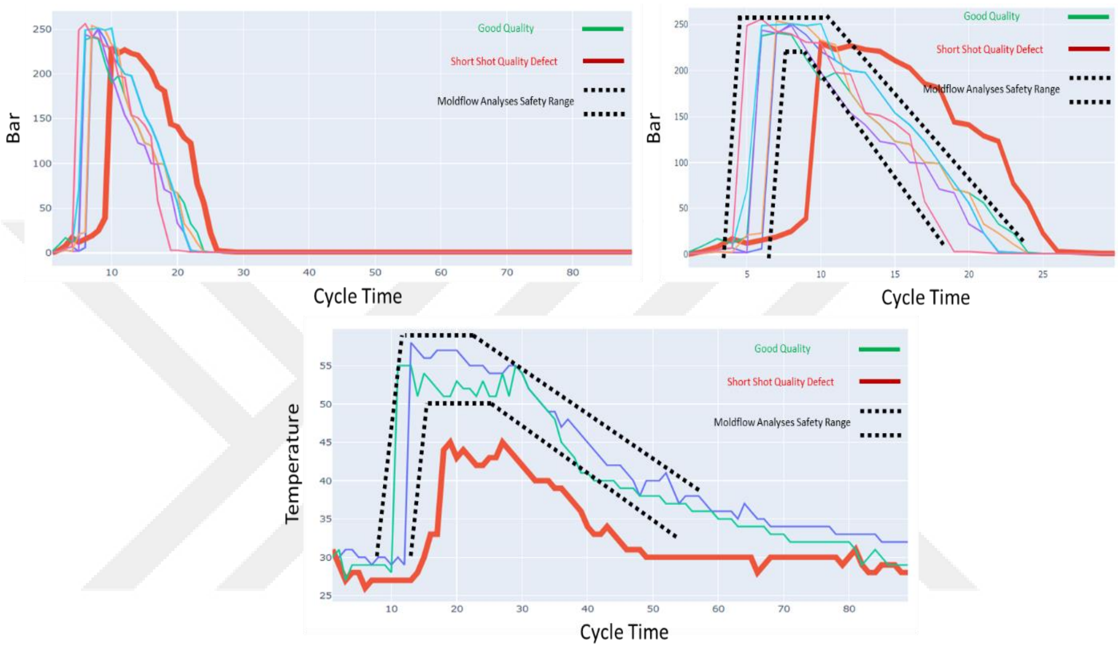

Figure 53: IMM Inmold Pressure and Temperature Time Series Data

Because of material fluctuations and different environmental conditions, often bad parts are produced at a constant machine setting.

This system controls the injection molding process fully automatically by adapting the machine parameters permanently. Because only this way it is guaranteed that molds can be transferred from one location to another. Ultimately, the quality parameters such as the compression, viscosity, or the molded part's dimension are controlled directly. The precondition is the separate measurement of cavity pressure and cavity temperature on the mold.

4.4 Summary

The application steps of the model to be developed are completed with the CRISP-DM method. As a result of these steps, the xgboost algorithm is used, which gives the best results according to performance metrics. Feature importances are removed, and the model is activated in production. The aim here is to understand short shot quality defect probability in advance in the production cycle and to tell the operator which parameter to change. Production is monitored in detail daily during the commissioning phase. In the one-month commissioning period, the short shot quality defect rate decreases from 2.45% to 1.73%. Production is monitored for three months, and the model is activated as an operator early warning system.

The development of this type of system takes approximately eight months in only one product variant. This situation is not sustainable for the industry. When modeling short shot quality defects, it is seen that the parameters in the IMM mold unit are particularly effective. However, in classical IMM machine controls, there is no system that measures melt plastic behavior. Therefore, it aims to strengthen the model with in-mold sensors to popularize this approach in the industry. Critical sensor positions are determined by mold flow analysis in production areas where MES infrastructure is not ready due to its structure. The safe pressure and temperature range required at these sensor positions are determined. According to the data read here, IMM feedback is given to the system to produce the ideal range in each cycle during production. This work starts in the first stage when the ML model is put into use. In the one-month commissioning period, the short shot quality defect rate decreases from 1.73% to 0.82%. This dramatic reduction results in an overall 80% improvement in short shot quality failure.

CHAPTER 5

SUGGESTIONS FOR FURTHER RESEARCH

The suggestions for improving the output in this study are as follows:

- The approach method in the created FEA model via mold flow analysis is Hele-Shaw. This approach not available for all plastic products due to its suitability for thin-walled parts. Therefore, it can be investigated and compared in other methods in the literature. A more accurate solution can be provided for potential short-shot areas by comparing other methods.

- The image processing algorithm created in the camera control station is established by finding the location of the template image from the large image with the template matching method. Performance metrics of the image processing algorithm can improve with different approaches in the literature.

- The data can be collected to include more process parameters from the auxiliary equipment in the process. In addition, time-series data such as pressure and temperature could not be obtained. New features can be added to improve the performance metrics of the predictive quality model developed within the scope of data mining.

- The plastic raw material MFI value could not be evaluated while modeling the short shot quality failure. This feature can be logged together with the environmental conditions to reveal the effect of raw material variability. It can be used to improve the performance metrics of the predictive quality model developed in this context.

- The different feature reduction techniques and imbalanced dataset approaches can be used while developing the predictive model. In addition, various machine learning techniques can be evaluated.

- At the last stage, the feedback system is developed with in-mold sensors. In the developed feedback system, the flow in the mold is balanced with specific process parameters with a high feature importance factor. It can be used to improve the predictive quality model's performance metrics with a full-scale data-mining study.

CHAPTER 6

DISCUSSIONS AND CONCLUSIONS

The FEA model analyzes the unexpected short shot quality defect seen during plastic parts production in injection machinery molding (IMM) via Autodesk Moldflow Analysis. The mold flow FEA model is verified in production. Thus the areas to be focused on are determined. IMM is integrated into MES and a quality control system with cameras installed at the end of the production line. The relationship between short shot quality defect and IMM critical process parameters extracted with data mining. The model is generated and commissioned in the production line. The different approach brought with in-mold sensors since integrating each line into MES and installing a camera measurement system can adversely affect industrial deployment sustainability. In each work step, the short shot quality defect is measured and recorded. The aim is to produce the system in the ideal range in each cycle by giving feedback to the IMM according to the in-mold sensors' data during production. In each work step, the short shot quality defect is measured and recorded. The main aim of this thesis is to avoid short-shot quality failure in plastic part production. The short-shot quality defect has not been completely eliminated in this study. However, it has decreased from 3.7% error rates to 0.82%. This dramatic reduction results in an overall 80% improvement in short shot quality failure. Conclusions and the contributions of the thesis are as following:

- It has been shown that FEA model using by Autodesk Molflow Analysis software can estimate the possible short shot quality defect area in the plastic part production.

- The short-shot quality defect failure decreased from 3.7% to 2.45% with a camera control measurement system installed at the end of the production line. In the scope of avoiding short-shot error, 32% improvement is achieved in terms of quality.

- The traceability function gained via RFID sensors placed on the pallets of the plastic parts. It has added value in terms of traceability in the supply chain process leading up to the customer.

- The developed models can be run on the SCADA PC of the machines in production. Thus, it can eliminate massive investment within the scope of digitalization.

- The short shot quality defect decreased from 2.45% to 2.32% with the dashboard of the process parameters prepared for the production team. It speeds up the process of gaining experience for operators by remote monitoring.

- The development of such a system takes about eight months in only one product variant of Tub part. This situation is not sustainable for the industry. When modeling short shot quality defects, the parameters in the IMM mold unit appear to be useful. However, there is no system that measures the melt plastic behavior in conventional IMM machine controls. For this reason, it has been considered to increase the applicability of the model with in-mold sensors in order for this approach to deployment in the sector. In production areas where MES infrastructure is not ready due to its structure, critical sensor positions are determined by mold flow analysis. The safe working range of the Pressure / Temperature sensors is clarified with the FEA model by using mold flow analysis. The in-mold sensors give feedback to IMM to realize the ideal range in every cycle during production. The short shot quality defect decreased from 1.73% to 0.82% by applying this approach. This dramatic reduction results in an overall 80% (%3.7:%0.82) improvement in short-shot quality failure.

BIBLIOGRAPHY

[1] Wang, X., Zhao, G., & Wang, G. (2013). Research on the reduction of sink mark and warpage of the molded part in rapid heat cycle molding process. Materials & Design, 47, 779–792.

[2] Kitayama, S., & Natsume, S. (2014). Multi-objective optimization of volume shrinkage and clamping force for plastic injection molding via sequential approximate optimization. Simulation Modelling Practice and Theory, 48, 35–44.

[3] Kitayama, S., Onuki, R., & Yamazaki, K. (2014). Warpage reduction with variable pressure profile in plastic injection molding via sequential approximate optimization. The International Journal of Advanced Manufacturing Technology, 72(5-8), 827–838.

[4] Moayyedian, M., Abhary, K., & Marian, R. (2015). New Design Feature of Mold in Injection Molding For Scrap Reduction. Procedia Manufacturing, 2, 241–245.

[5] Zhou, H. (2013) Computer Modeling for Injection Molding: Simulation, Optimization, and Control; Wiley: Hoboken, NJ.

[6] Kumar, A.; Ghoshdastidar, P.S.; Muju, M.K. (2002) Computer simulation of transport process during injection mold-filling and optimization of the molding conditions. Journal of Materials Processing Technology, 120, 438–449.

[7] Courbebaisse, G. (2005) Numerical simulation of injection molding process and the pre-modeling concept. Computational Material Science, 34, 397–405.

[8] Fernandes, C., Pontes, A. J., Viana, J. C., & Gaspar-Cunha, A. (2016). Modeling and Optimization of the Injection-Molding Process: A Review. Advances in Polymer Technology, 37(2), 429–449.

[9] Jones D, Snider C, Nassehi A, Yon J, Hicks B (2020) Characterizing the Digital Twin: A systematic literature review, CIRP Journal of Manufacturing Science and Technology 29(A): 36-52

[10] Grieves M, Vickers J (2017) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems, Transdisciplinary perspectives on complex systems, 85–113.

[11] Lohtander M, Ahonen N, Lanz M, Ratava J, Kaakkunen J (2018) Micro manufacturing unit and the corresponding 3D-model for the digital twin. Procedia Manuf 25:55–61.

[12] Kritzinger W, Karner M, Traar G, Henjes J, Sihn W (2018) Digital twin in manufacturing: a categorical literature review and classification, IFAC-PapersOnLine 51:1016–102

[13] Zhuang C, Liu J, Xiong H (2018) Digital twin-based smart production management and control framework for the complex product assembly shopfloor, Int J Adv Manuf Technol 96:1149–1163.

[14] Gartner (2017) Prepare for the Impact of Digital Twins, September 18, 2017 (Contributor: C. Pettey) https://www.gartner.com/smarterwithgartner/prepare-for-the-impact-of-digital-twins/, Accessed 18/05/2020.

[15] Mahnke, W., Leitner, S.-H., & Damm, M. (2009). OPC Unified Architecture.

[16] OPC UA interfaces for plastics and rubber machinery,(2019).Data exchange between injection moulding machines and MES

[17] Teng, S. Y., Touš, M., Leong, W. D., How, B. S., Lam, H. L., & Máša, V. (2020). Recent advances on industrial data-driven energy savings: Digital twins and infrastructures. Renewable and Sustainable Energy

[18] Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. Big Data & Society, 5(2), 205395171878631.

[19] Popp, J., Tuchin, V., Chiou, A., Heinmann, S. (2012). Handbook of Biophotonic. Weinheim: WileyVCH Verlag GmbH & Co. KGaA.

[20] Heleno, P., Davies, R., Correia, B., Dinis, J. (2001). A Machine Vision Quality Control System for Industrial Acrylic Fibre Production. EURASIP Journal on Applied Signal Processing, 7(1), 728-735.

[21] Amza, C. G., & Cicic, D. T. (2015). Industrial image processing using fuzzy-logic. Procedia Engineering, 100, 492-498

[22] Szelisky, R. (2011). Computer Vision. London: Springer.

[23] Mohan, A., & Poobal, S. (2017). Crack detection using image processing: A critical review and analysis. Alexandria Engineering Journal.

[24] Datta, A., Kim, J.-S., & Kanade, T. (2009). Accurate camera calibration using iterative refinement of control points. 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops.

[25] Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., … Yu, T. (2014). scikit-image: image processing in Python. PeerJ, 2, e453.

[26] Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216–231.

[27] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1).

[28] Kotarba M (2018) Digital Transformation of Business Models Foundations of Management 10: 123-142

[29] Schallmo D, Williams C (2018) Digital Transformation of Business Models-Best Practices, Enablers and Roadmap, XXVIII ISPIM Innovation Conference Austria Vienna on 18-21 June 2017

[30] Demartini, M., Evans, S. and Tonelli, F., 2019. Digitalization Technologies for Industrial Sustainability, Procedia Manufacturing, 33, 264-271.

[31] He, B., Bai, KJ. Digital twin-based sustainable intelligent manufacturing: a review. Adv. Manuf. (2020).

[32] Wildhirt K, Seidel C, Bub U, Josten M, Kühr S (2019) Digitalization Partnership: How GKN Established a Digital Platform with 3YD to Realize the Disruptive Potential of Metal Additive Manufacturing. In: Urbach N., Röglinger M. (eds) Digitalization Cases Management for Professionals.

Springer, Cham.

[33] Weiblen T, Chesbrough HW (2015) Engaging with startups to enhance corporate innovation. Calif Manage Rev 57(2):66–9

[34] Kashyap, S., & Datta, D. (2015). Process parameter optimization of plastic injection molding: a review. International Journal of Plastics Technology, 19(1), 1–18.

[35] Dang, X.-P. (2014). General frameworks for optimization of plastic injection molding process parameters. Simulation Modelling Practice and Theory, 41, 15–27

[36] Chen, W.-C., Fu, G.-L., Tai, P.-H., & Deng, W.-J. (2009). Process parameter optimization for MIMO plastic injection molding via soft computing. Expert Systems with Applications, 36(2), 1114–1122.

[37] Fabian Pedregosa; Gaël Varoquaux; Alexandre Gramfort; Vincent Michel; Bertrand Thirion; Olivier Grisel; Mathieu Blondel; Peter Prettenhofer; Ron Weiss; Vincent Dubourg; Jake Vanderplas; Alexandre Passos; David Cournapeau; Matthieu Perrot; Édouard Duchesnay (2011). "Scikit-learn: Machine Learning in Python". Journal of Machine Learning Research.

[38] Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP, 79, 403–408.

[39] Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. Information Fusion, 37, 132–156.

[40] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453.

[41] Adler, J., & Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry Part A, 77A(8), 733–742.

[42] Feng, W., Zhu, Q., Zhuang, J., & Yu, S. (2018). An expert recommendation algorithm based on Pearson correlation coefficient and FP-growth. Cluster Computing.

[43] Bro, R., & Smilde, A. K. (2014). Principal component analysis. Anal. Methods, 6(9), 2812–2831

[44] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.

[45] Le, T., Hoang Son, L., Vo, M., Lee, M., & Baik, S. (2018). A Cluster-Based Boosting Algorithm for Bankruptcy Prediction in a Highly Imbalanced Dataset. Symmetry, 10(7), 250.

[46] Kamiran, F., & Calders, T. (2011). Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1), 1–33.

[47] Yin, Z., Wang, Y., Liu, L., Zhang, W., & Zhang, J. (2017). Cross-Subject EEG

Feature Selection for Emotion Recognition Using Transfer Recursive Feature Elimination. Frontiers in Neurorobotics, 11.

[48] Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magnetic Resonance in Medicine, 62(6), 1609–1618.