Dissertation on

"Cost Estimation for Aerospace Manufacturing using Machine Learning models"

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

UE21CS320A – Capstone Project Phase - 1

Submitted by:

Deepika Indran	**PES2UG21CS158**
Devansh Guttikonda	**PES2UG21CS160**
Lakshmi Narayanan	**PES2UG21CS251**
Tanishka Pasarad	**PES2UG21CS570**

Under the guidance of

Dr. Sandesh BJ
Professor & Chairperson, Computer Science & Engineering
PES University

January - May 2024

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

‘Cost Estimation for Aerospace Manufacturing using Machine Learning Models’

is a bonafide work carried out by

Deepika Indran	**PES2UG21CS158**
Devansh Guttikonda	**PES2UG21CS160**
Lakshmi Narayanan	**PES2UG21CS251**
Tanishka Pasarad	**PES2UG21CS570**

In partial fulfillment for the completion of sixth-semester Capstone Project Phase - 1 (UE21CS320A) in the Program of Study -Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Jan. 2024 – May. 2024. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 6th-semester academic requirements in respect of project work.

Signature	Signature	Signature
Dr. Sandesh B J	Dr. Sandesh B J	Dr. B K Keshavan
Chairperson	Chairperson	Dean of Faculty

External Viva

Name of the Examiners	**Signature with Date**
1. ____________________________	____________________________
2. ____________________________	____________________________

ACKNOWLEDGEMENT

I would like to express my gratitude to **Dr.Sandesh BJ**, Professor & Chairperson, Department of Computer Science and Engineering, PES University, for his/ her continuous guidance, assistance, and encouragement throughout the development of this UE21CS390A - Capstone Project Phase – 1.

I am grateful to all Capstone Project Coordinators, for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Sandesh B J, Professor & Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro-Chancellor, PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University, and Prof. Nagarjuna Sadineni, Pro-Vice Chancellor, PES University, for providing me with various opportunities and enlightenment every step of the way. Finally, Phase 1 of the project could not have been completed without the continual support and encouragement I have received from my family and friends.

DECLARATION

We hereby declare that the Capstone Project Phase - 1 entitled **“Cost Estimation for Aerospace Manufacturing using Machine Learning models”** has been carried out by us under the guidance of **Dr. Sandesh BJ, Chairperson , Computer Science and Engineering** and submitted in partial fulfillment of the course requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester January – May 2024. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES2UG21CS158	**Deepika Indran**
PES2UG21CS160	**Devansh Guttikonda**
PES2UG21CS251	**Lakshmi Narayanan**
PES2UG21CS570	**Tanishka Pasarad**

ABSTRACT

Our project aims to build a cost estimation tool for aerospace manufacturing. This industry is highly complex and multidimensional because it uses lots of different materials and has intricate processes. The current cost prediction models are not very accurate because they don't look at the historical data of past projects closely enough. Our goal is to use Machine Learning to make a self-learning and more efficient model that learns from past projects and understands the aerospace industry better.

Our model will look at attributes such as the type of materials used, the size of the parts, and how cost drivers are identified to predict costs. The cost estimation tool will be well-equipped to use current market information to ensure the predictions remain accurate with any fluctuations in prices, machining or labor costs. This will facilitate better identification of key cost drivers and improve the overall performance of the tool. Another important feature of our tool is that it can adapt to different project details. For example, if a project needs a specific cutting tool, the model will adjust the cost estimate accordingly. This helps the cost estimation tool to cater to a wide variety of project requirements across the various use cases that are present in the aerospace industry. We will keep updating our model to make sure it maintains its accuracy in predicting costs.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The aerospace manufacturing sector is a vast division within which accurate cost estimation plays a significant role. Accurate cost estimation proves vital for making informed decisions,optimization of resources as well as to maintain competitiveness within this vast and dynamic sector

Currently cost estimation is being done manually in many of the companies within this industry, that proves to be extremely time consuming and also quite prone to errors which result in less accurate estimations

Cost estimation in the aerospace industry needs to factor in many different features such as raw material costs,labor rates,machining expenses and overhead costs which some traditional methods fail to be precise as well as encompass all these factors and be adaptable to changes since some of the features used are dynamic in nature

To address these concerns, our project is developing a customized cost estimation model specifically for the aerospace manufacturing sector. We aim to close the gap left out by the pre existing methods by using robust machine learning algorithms,precise datasets, and insights from industry experts to create an adaptable and accurate solution.

In the end we wish to make use of the data-driven insights as well as predictive analytics to estimate cost efficiently in the aerospace manufacturing sector which will in turn help support strategic decision-making throughout production.

CHAPTER 2

PROBLEM STATEMENT

Most companies in the aerospace manufacturing industry currently tend to rely upon manual cost estimation practices.They will first and foremost rely on using conventional historical data as well as generic concepts which do not help accurately capture the complex nuances of each and every project.Then we also have the time and energy consuming portion of manual cost estimation which are the manual audits.Third,most of the times manual methods are quite prone to errors and inconsistencies which in turn greatly affects the predictions being made.Lastly,manual estimates struggle to integrate technological developments as well as fluctuating market prices which limit its adaptability in this sector

Now, looking at the limitations of the manual cost estimation practices, we know that there is a need for an effective and automated solution. Our solution will address these challenges of accuracy, efficiency, reliability and overcoming changes in manual processes.Our goal through this capstone project is to develop and provide companies in the aerospace industry with a tool that provides accurate cost estimates for the aircraft components being manufactured so that they decision making can be improved,resources optimised and competitiveness maintained

To create this tool for the aerospace component manufacturing companies we will be making use of an abundant amount of historical cost data which contains information on material properties,manufacturing processes,design specifications and many other external features.
The machine learning algorithms will be able to learn the complex patterns and relationships which are used to estimate costs and do it with greater accuracy than the methods that are currently being employed .By using machine learning models we will also be able to capture the ever evolving industry dynamics as well as help reduce burden of work on the employees,streamline their workflow,minimise the errors and overall optimise the whole process of cost estimation

CHAPTER 3

LITERATURE SURVEY

PAPER 1-

Title: A parametric-fuzzy logic system for paint cost estimation of products with uncertain geometric characteristics

Objective:

- The study's aim is to build a framework for cost estimation of paint cost in spraying techniques for wooden artifacts based on fuzzy theory.
- The study wanted to showcase the dependency of the paint loss on the parameters of the wooden artifacts and not defining the loss as a constant function based on surface area.

Working:

- The study involved in the creation of a fuzzy set with the parameters surface area, volume, and projected area. Fuzzy rules were then generated and tweaked until results were satisfactory.
- Mamdani's fuzzy inference method was then employed to obtain an output set which was the amount of paint required.These outputs were then aggregated followed by defuzzification to generate the numerical paint estimate value.

Advantages:

- Predicted results from the fuzzy model as well as a traditional parametric cost function were compared and the fuzzy model was superior. It provided more accurate results especially for larger products in an upright placement on the spraying rack.
- Provides a detailed framework and structure for building a cost-fuzzy system.

Disadvantages:

- Although the study looks into the parameters of the products, it does not consider the complexity in design of each artifact.

- The study focused primarily on paint cost estimation and not the other factors of aerospace cost estimation such as raw materials, bought-out-parts, machining.

Result:

- The study revealed an RMSE of 5.66 for traditional parametric models and 0.96 for fuzzy inference system models.

PAPER 2-

Title: Product Cost Estimation Using Fuzzy Logic

Objective:

- This study proposes a cost estimation model for producing a pressure vessel using Fuzzy Inference Engines.
- It also provides a comparative study between a second order regression model and fuzzy logic for product cost estimation.

Working:

- The regression model uses a special equation for each case using a subset of variables that are most influential to the cost but are not correlated. The parameters in the equation are determined through least square regression.
- In the case of a fuzzy inference engine, first a membership function is defined for each fuzzy set and is applied onto the input parameters to determine truth values. The truth values for each rule are then computed and applied to the conclusion part of each rule.
- The final step is defuzzification of the fuzzy output to crisp numerical values.

Advantages:

- Provides the advantages and limitations of both second order regression and fuzzy systems in a cost estimation context
- Fuzzy logic thrives when historical data is not as accurate and imprecise as well as cost prediction during early stages of design.

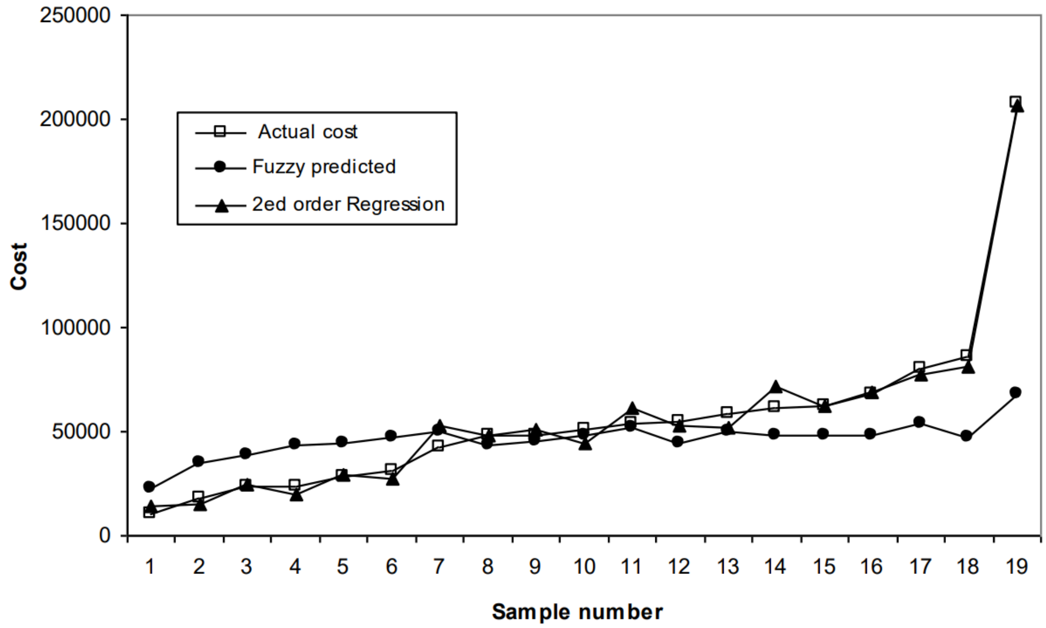
Disadvantages:

- The study takes in specific test cases to generate it’s results and is not a large sample of predictions for validation.

- The R2 value of second order regression was 99.65% which may indicate potential overfitting of the model

Results:

- The R2 value of second order regression was 99.65% whereas the R2 value of the fuzzy model was 67%



-

Figure 1. Comparison between Fuzzy Logic and Regression

PAPER 3-

Title: A hybrid software cost estimation approach utilizing decision trees and fuzzy logic

Objective:

- The study aims to explore software cost estimation models using fuzzy implication systems.

Working:

- The proposed fuzzy cost estimation system is carried out in the following stages
  - (i) Pre-processing of the cost driver data
  - (ii) Creation and evaluation of FDT
  - (iii) Implementation of the FIS
  - (iv) Effort Estimation

Advantages:

- The novelty aspect of the following paper is in the generation of robust fuzzy decision trees, with high significance and classification rates.
- The study gives a base knowledge for other cost estimation models that have been used for software cost estimation and a comparative study between the model developed and traditional models.

Disadvantages:

- Software cost estimation and its parameters can not be directly compared to product cost estimation in a manufacturing context as their parameters vary.

Result:

- 3 publicly available datasets were used for validation, accurate prediction costs were achieved for ISBSG and DESHARNAIS datasets.

PAPER 4-

Title: Manufacturing cost prediction through data mining

Objective:

- The study explores the feasibility of predicting manufacturing costs of a product with non-parametric methods through data mining.
- Utilizes a dataset with categorical and numerical values to make the prediction.

Working:

- K-medoids clustering algorithm is employed to cluster similar products together.
- Comparison of products is done by calculating mathematical similarities which is then used for clustering.
- Cost estimation is carried out by applying regression techniques on the clusters generated.
- New products can be assigned to pre-established clusters.

Advantages:

- This framework handles datasets with categorical and numerical values which is in a similar format our cost estimation model is being applied on.
- The parameters or the regression technique for each cluster varying based on the nature of the cluster generating more accurate results

Disadvantages:

- Clusters could potential destroy product families and product hierarchy of the data.

Results:

- The regression models had a 5% confidence interval in their generated estimates

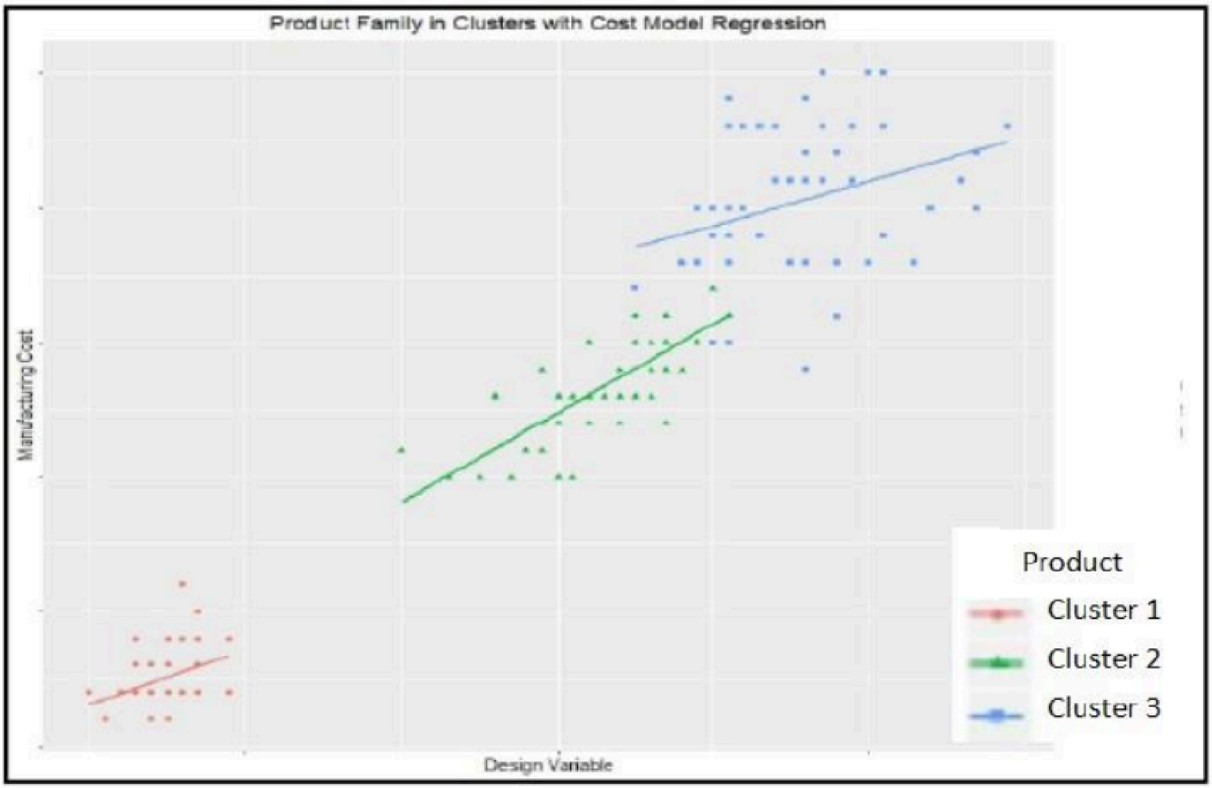


Figure 2. K Medoids clustering to create product families

CHAPTER 4

DATA

4.1. Overview

We have received a detailed "Part Requirement Sheet" from a reputed company in the Aerospace industry. The following columns of the dataset are of utmost importance-

- Quantity- The quantity of products estimated in terms of the annual units that can be produced.
- Form- The form of the products can range from round (abbreviated as RND) and plate to bar and sheet.
- Alloy, Temper and Specifications- These are very specific requirements according to the customer requirements and material type.
- Part size in inches- The height, width and length of the product are defined in inches
- Est. RM size- The height, width and length of the raw material are defined in inches
- Machining- Under machining, the machine type and operation related specifications are mentioned
- Surface treatment- Post machining, surface treatments like anodising and galvanising come into play.
- Assembly- The last cost that has to be taken into consideration is the assembly cost and the need for any special cutting tools or bought out parts that have to be sourced.

Quantity			Lead Time in Weeks			Raw-material													Machining			
											Part size in inches			Est. RM Size in								
SS/AC	EAU	MBQ	RM	Manfg.	Total	Form	Alloy	Temper	Specification	Specific / Alternate	Thk/ W thk	Width/ Dia	Length / GF	Thk/ W thk	Width/ Dia	Length / GF	No.of parts / block	Potential Source	Machine	Setup Time in Hrs	Cycle Time in Hrs	Bench Work Time in Hrs
-	350	50	16	8	24	RND	6061	T6	QQ-A-200/8	T651, QQ-A-225/8		2.94	1.96		3.25	2.36	1	Sankap/TW	MT5A0.2D	2.00	0.50	0.10
-	1310	200	16	8	24	RND	6061	T6	ASTM B221			2.50	0.70		2.75	1.10	1	Sankap/TW	TM4A0.1D	2.00	0.50	0.10

Figure 3. Overview of Dataset (part 1)

Machining				Inspection time		Surface Treatment		Bought-out			Assembly		Estimated Base Cost/Unit in USD						
Machine	Setup Time in Hrs	Cycle Time in Hrs	Bench Work Time in Hrs	CMM (Hrs/batch)	Mannual (Hrs/batch)	Process	Potent ial Source	PN's / Assy.	Total Qty./Assy.	Potenti al Source	Setup time in Hrs	Process time in Min's	Fixture Manfg.	Special C.Tools	Cutting Tools	Assembly Tools	Raw Material	Bought out	Special Process
MT5A0.2D	2.00	0.50	0.10	1.25	4.50	CCC+AN	VMF	-	-	-	-	-	750.00	-	2.50	-	12.56	-	4.16
TM4A0.1D	2.00	0.50	0.10	5.00	18.00	AN	VMF	-	-	-	-	-	562.50	-	2.50	-	4.20	-	1.44

Figure 4. Overview of Dataset (part 2)

The part requirement sheet serves as the base for cost estimation. It serves as a blueprint for understanding the intricacies of the overall cost estimation process. Along with the available data columns, we will be augmenting additional columns such as weight per part. In our pre-processing, we will augment additional data rows according to the different permutation and combination specifications given to us by industry experts in order to create a more comprehensive dataset. The raw dataset has to be categorised and normalised before we proceed with the model development.

RM Form	Material	RM Alloy	RM temper	RM Specification
Plate	Aluminium	2011	T3	BMS7-371
Round	Steel	2024	T35	AMS5659 TYPE I

Figure 5. Overview of the dataset that maps raw materials and their specifications

Anodising				Passivation				Primer				Top Coat					Silver Plating			
Spec	Type	Class	Color	Spec	Type	Class	Method	Spec	Type	Class	Grade	Spec	Type	Class	Grade	Color code	Spec	Type	Class	Grade
MIL-A-8625	1	1	No Dye	AMS2700	1	1	1	MIL-PRF-23377	1	C1	A	MIL-PRF-85285	1	H	N	Red	QQ-S-365	1		A
CPS1000-01	1B	1C	Black	BAC5625	2	2	2	BMS10-11	2	C2	B	BMS10-60	2	NH	P	Orange-Yellow		2		B

Figure 6. Overview of the various special processes that can be applied post-machining

CHAPTER 5

Software Requirements Specifications

5.1 Introduction

The purpose of this SRS document is to delineate the requirements for a software system tailored for aerospace manufacturing cost estimation. The software will utilize cutting-edge technology to simulate manufacturing processes, offering comprehensive analysis, customized cost projections, and strategies for optimizing resource utilization and minimizing expenses.

5.2 Literature Survey on Existing Systems

A. P. Chansaad, S. Chaiprapat and P. Yenradee(2013). built a framework for cost estimation of paint cost in spraying techniques for wooden artifacts based on fuzzy theory. The study wanted to showcase the dependency of the paint loss on the parameters of the wooden artifacts and not defining the loss as a constant function based on surface area. Predicted results from the fuzzy model as well as a traditional parametric cost function were compared and the fuzzy model was superior. It provided more accurate results especially for larger products in an upright placement on the spraying rack. [1]

Baioumy, S(2010). proposed a cost estimation model for producing a pressure vessel using Fuzzy Inference Engines. It also provides a comparative study between a second order regression model and fuzzy logic for product cost estimation. The regression model uses a special equation for each case using a subset of variables that are most influential to the cost but are not correlated. In the case of a fuzzy inference engine, first a membership function is defined for each fuzzy set and is applied onto the input parameters to determine truth values. The truth values for each rule are then computed and applied to the conclusion part of each rule. These truth values are then defuzzified to generate the cost estimates. [2]

Papatheocharous E, Andreou AS(2012). used a novel approach for cost estimation with the generation of robust fuzzy decision trees, with high significance and classification rates.

The study also gives the base knowledge for other cost estimation models that have been used for software cost estimation and a comparative study between the model developed and traditional models. [3]

Díaz, Andrea, et al (2020). Explores the prediction of product manufacturing costs with non-parametric methods through data mining. It utilizes a dataset with categorical and numerical values to make the prediction. A k-medoids clustering approach is taken to cluster similar products together, regression techniques are then applied onto each cluster to generate cost estimates. [4]

5.3 Product Perspective

The Aerospace Cost Estimation Tool represents a breakthrough solution addressing the challenges inherent in conventional cost estimation practices within the industry. The main idea is to create a user-friendly tool with intuitive interfaces, harnessing the power of machine learning that can offer near precise cost projections, streamline resource allocation and enhance decision making for stakeholders by providing them a competitive edge over other bidders in the bidding process.

5.3.1. Product Features

The major/significant features of the product are as follows:

- Cost Modelling- Enables users to simulate aerospace manufacturing cost estimation process by leveraging machine learning techniques.

- Dashboard- A dashboard with an intuitive interface where cost estimators can view the projections and the overall process followed in estimating costs.

- Report Generation- A comprehensive report can be downloaded which can be shared with other stakeholders.

5.3.2. User Classes and Characteristics

The various user classes using the produce are as follows:

Aerospace Cost Estimators:

- These are experts that estimate the cost of manufacturing in the aerospace industry.
- Use this tool to carry out a whole cost estimate relying on inputs such as project particulars, materials, processes and labour demands.
- Use the dashboard for visualization of estimation process, review of graphs and adjustment of input values for recalibration and optimization purposes.

Administrators:

- They oversee the operational aspects of cost estimating tool.
- Manage data entry, preprocessing and maintain a database containing relevant costing factors and standards within the industry in question.

- Ensure security protocols are adhered to while conducting routine maintenance activities aimed at upholding integrity and precision levels of the tool.

5.3.3. Operating Environment

The tool is built using the MERN (MongoDB, Express.js, React.js, Node.js) stack. The tool should also be integrated with existing company software.

5.3.4. General Constraints, Assumptions and Dependencies

Dependencies:

- The tool should adhere to data protection laws and ensure adequate privacy.
- A cross-compatible tool that can work seamlessly with a stable internet connection.

Assumptions:

- The data input must not contain any errors and must be of high quality.
- Users are expected to have a basic understanding of the aerospace manufacturing industry and the various processes involved in the cost estimation process in order to interpret results effectively.

Risks:

- Hardware malfunctions may cause a loss in accuracy or disrupt the cost estimation process.
- Any software bugs in the application can lead to faulty estimates.

5.4 Functional Requirements:

- A user-friendly interface is of utmost importance to ensure that the cost estimation team can use it easily.
- The input parameters should include material dimensions, shape, specifications, temper conditions, and alloy requirements.
- The input parameters should be modifiable. There should also be admin access to modify certain parameters uniformly for all users.
- Error handling should be smooth with appropriate error messages displayed.
- The model is expected to learn the training dataset well and capture all the intricacies of the complex dataset. The whole process of cost estimation is quite dynamic and the model is expected to accommodate the changes in input parameters and provide near-accurate results.
- Use of visuals like charts, graphs, and tables for presenting cost breakdowns and comparisons between them.
- Finally, there should be an option to generate a report which can be shared by the cost estimation team with other stakeholders.

5.5 External Interface Requirements:

User Interfaces

- There should be input boxes and drop-down menus to specify input parameters.
- The tool must have a user-friendly dashboard to view the results of the model and the various visualizations.
- The users should be able to customize or change certain input variables based on results and the model should recaliberate.

Software Requirements:

- Deep learning frameworks such as sci-kit learn for model building.
- The software should be compatible with existing programs used in aerospace manufacturing.

5.6. Non-Functional Requirements:

Performance Requirement:

- The system should support concurrent users.
- The system should meet the GPU requirements of the tool.
- The tool must have a minimal response time.

Safety and Security Requirements:

- The integrity of the cost estimation data must be maintained.
- Clear error messages should pop up on any system failure.
- Confidentiality must be maintained for all customer/client related data.
- The system must adhere to security standards and regulations as practiced in the aerospace industry.
- The system should validate the user/admin before giving access to the tool and data.

CHAPTER 6

SYSTEM DESIGN

6.1 Existing systems

In aerospace manufacturing, estimating costs has traditionally been a manual and tedious process, where established and pre-defined formulas are used adhering to industry standards, to calculate expenses like material costs, labor hours, machining expenses, and overhead costs. Human errors and misclassification of requirements were prevalent in this method of cost estimation. A major disadvantage of the existing system is that it does not incorporate historical data of previous projects or real-time market trends which may lead to ignoring the complex nature of aerospace projects and further missing out on important factors that affect the cost estimate.

Machine learning is being explored to improve cost estimation in aerospace manufacturing, but it's still in the early stages. The current methods have some issues as they don't make good use of past data, they can't adjust to real-time changes in data, and they rely heavily on manual inputs, which can lead to errors. These challenges make it difficult to accurately compare costs, validate estimates, and predict future expenses in the aerospace industry.

6.2 Design Considerations:

The proposed system using machine learning promises a solution for the shortcomings of the existing system. Through its implementation, the cost estimation tool offers many benefits, which include increased accuracy, usability, real-time data integration, security, privacy, speed, and performance. These advantages prove that it could possibly be beneficial to implement machine learning for cost estimation in the aerospace manufacturing industry. The following are the advantages of the proposed cost estimation tool:

1) It improves accuracy and makes the tool adaptable as it provides precise estimations by integrating machine learning algorithms, various types of input parameters, and historical data to account for industry variations.

2) The user interface and experience are intuitive as it contains features that are easy to locate and allows the users to easily input parameters and navigate the system effortlessly. Multiple fields for different formats of input are provided for the user to provide as many details as possible.

3) Real-time data is captured so that dynamic adjustments can be made to the manufacturing costs based on the changes in manufacturing material, labour or machine prices. This makes the cost estimations accurate irrespective of the product specification or requirement.

4) The system displays clear visualizations of charts, graphs, and tables to present cost breakdowns and comparisons in an understandable manner.

The proposed system prioritizes high availability, speed, and performance. Continuous monitoring and maintenance ensure high efficiency and performance while the machine learning models and high-performance computing resources ensure that the cost estimations will be prompt.

6.3 Architecture

Chosen Architecture: MVC (Model View Controller):
We want to implement the Model-View-Controller (MVC) architecture for our cost estimation tool, while using machine learning. The following are the advantages of using the MVC architecture for our proposed system

- MVC separates the application logic into three distinct components - Model, View, and Controller. This organizes the functionalities, making it easier to maintain
- Each component has the facility to progress independently and be modified easily. This feature is useful for scaling and adapting to the various requirements and product specifications that come from different projects across the globe.
- The modular components of MVC can be reused in the cost estimation tool. This improves the quality of the code and program structure.
- MVC supports the addition of new features and functionalities without significantly affecting existing code, making it suitable for projects with evolving requirements.
- It enables parallel development of different components by different members, fostering efficient collaboration and ensuring the overall robustness of the application.

Model

The Model represents the domain-specific data and the business logic of the application. In our cost estimation tool, the following components would constitute the Model:

- Data Preprocessing Module: preprocesses the input data, including feature scaling, encoding categorical variables, handling missing values and preparing it for the model.
- (Fuzzy Decision Tree Model) or (Clustering + Regression Model): represents the machine learning model responsible for estimating costs based on the input parameters.

View

The View is the user interface components responsible for displaying information to the user and taking in user inputs. In our cost estimation tool, the View consists of the following components:

- Dashboard Interface: Provides a user-friendly interface for users to input parameters related to the part being manufactured and view the estimation results. It includes forms, input fields, charts, graphs, and tables.
- Visualization Module: Generates the visualizations of the cost estimation results to present the cost breakdowns and comparisons to the user.

Controller

The Controller acts as an intermediary between the Model and the View, handling user inputs, updating the Model, and updating the View accordingly. In our cost estimation tool, the Controller consists of the following components:

- Input Handling Module: It ensures that the input parameters are correctly formatted and valid before being passed to the machine learning model to produce the cost estimate.
- Estimation Controller: It is responsible for the process of cost estimation by calling and allocating the appropriate methods from the Model and updating the View with the estimation results.

6.4 HLD (High Level Design) :

We are exploring a range of machine learning models, guided by our literature review out of which the one with the best accuracy will be selected. The 2 key machine learning models which have proven to be highly effective for use cases similar to ours are the fuzzy decision trees and a hybrid model of clustering and regression.

Fuzzy Decision Tree:

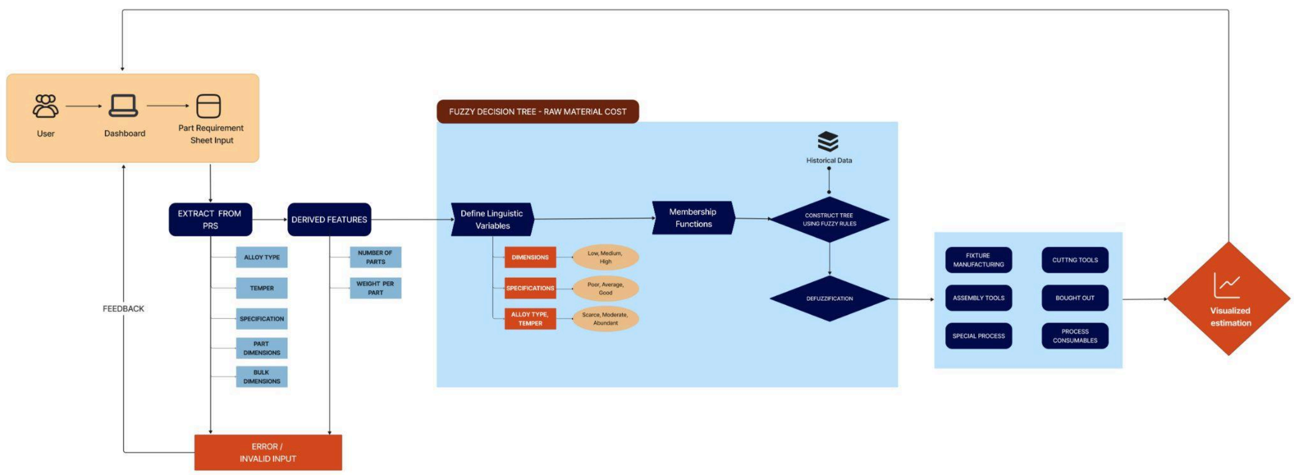


Figure 7. Architecture Diagram for Fuzzy Decision Tree

Hybrid of Clustering and Regression:

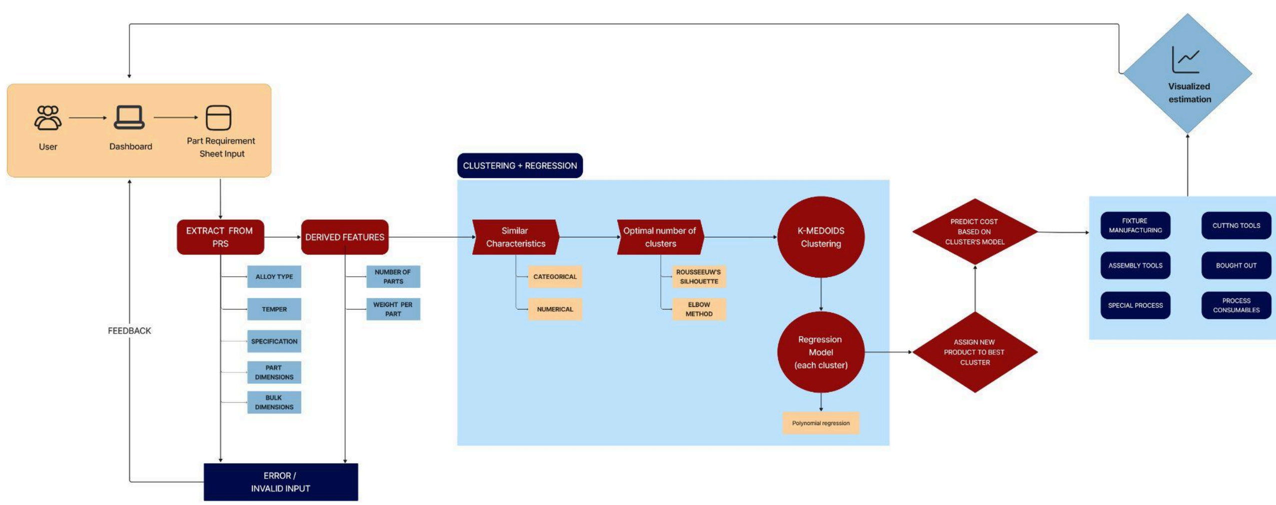


Figure 8. Architecture Diagram for Clustering and Regression

6.5 Design Description:

Our cost estimation model for the aerospace manufacturing industry consists of three main modules:

1. User Interface
2. Cost Estimator
3. Database

User Interface

This module allows users to interact with the system. It provides functionalities to the following:

1. Enter parameters relevant to the aerospace manufacturing project (enterParameters())
2. Display the estimated cost of the project (displayCostEstimate())

Cost Estimator

This module is the core of the system and performs the cost estimation. It interacts with the other two modules as follows:

1. Retrieves historical data from the database (retrieveHistoricalData())
2. Analyzes the data entered by the user through the user interface (analyzeInputData())
3. Performs the cost estimation using the historical data and user-provided data (performCostEstimation())
4. Stores the cost estimation result in the database (storeCostEstimationResult())

Database

This module stores the historical cost data used by the cost estimator. It interacts with the cost estimator module as follows:

1. Provides historical data when requested by the cost estimator (retrieveData())
2. Stores the cost estimation results sent by the cost estimator (storeData())

ER DIAGRAM:

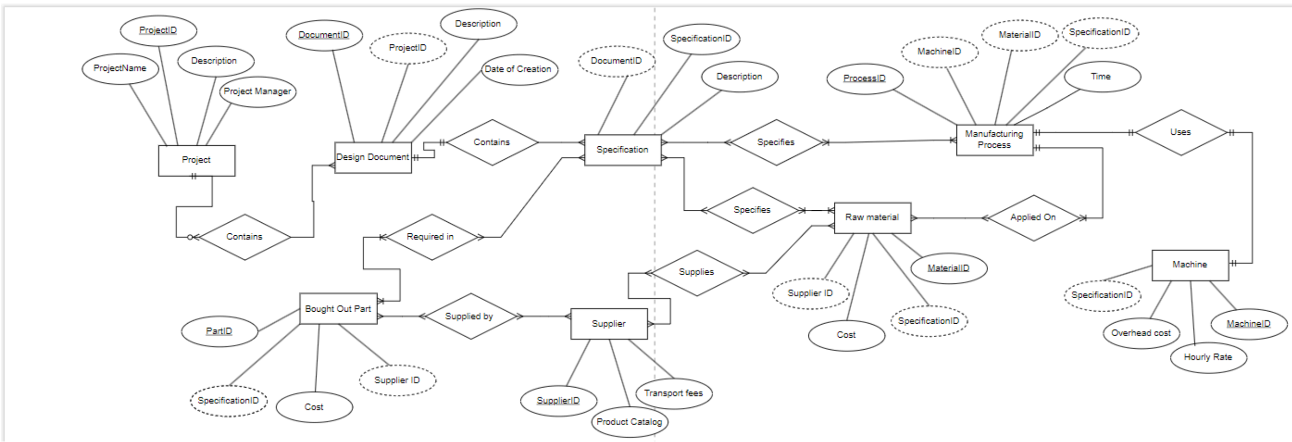


Figure 9. ER Diagram for Cost Estimation

#	Entity	Name	Definition	Type
ENTITIES				
1.	Project	Project Name	Aerospace manufacturing process	Document
2.	Design Document	Document	Design of the manufactured components in a project	Document
3.	Raw Material	Material	Operations are performed on raw materials in house for manufacturing products	Physical
4.	Bought out parts	Parts	Parts sourced from external	Physical

			suppliers and not manufactured	
5.	Manufacturing Process	Process	Special techniques used to manufacture a product	Physical
6.	Machine	Machine	Processes are carried out using machines	Physical
7.	Supplier	Vendor	Sells goods to that are needed in manufacturing process	Service
#	**Attribute**	**Name**	**Definition**	**Type (size)**
DATA ELEMENTS				
1.	Manufacturing Process	Time	Time to execute a process	DateTime
2.	Machine	Hourly Rate	Cost of using machine per hour	Float
3.	Raw materials	Cost	Cost of procuring raw materials	float
4.	Project	Description	Details description of the project	Str
5.	Machine	Overhead Cost	Cost of procuring machine	float
6.	Supplier	Transport fees	Cost of transporting goods in bulk orders	float

Table 1. Entity Attributes Table

MASTER CLASS DIAGRAM :

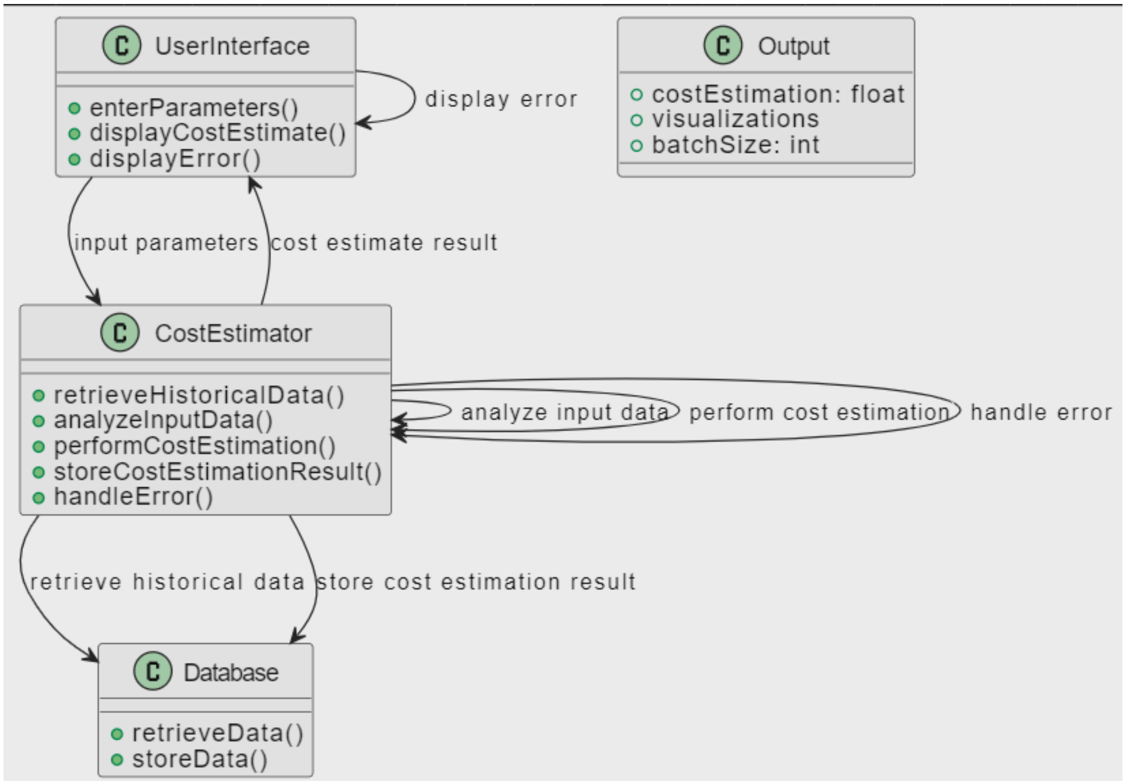


Figure 10. Master Class Diagram

Reusability Considerations:

This project can benefit from reusability at various stages.

Utilizing Existing Libraries and Frameworks:

1. Deep Learning Libraries: Libraries like TensorFlow or PyTorch can be leveraged for building and deploying the deep learning models used for cost estimation. These libraries offer pre-built functionalities for common tasks like data manipulation, model building, and training, saving development time.

2. Data Preprocessing Tools: Frameworks like Pandas or Scikit-learn provide functionalities for data cleaning, normalization, and feature engineering, which are crucial steps before feeding data into the deep learning models. These can be reused across different projects requiring similar data preparation tasks.

Building Reusable Components within the Project:

Cost Estimation Engine: The entire process of cost estimation, which includes retrieving data, analyzing the user input, and performing the necessary calculations, can be considered as a reusable module. It is implemented as a generic module that adapts to the different types of aerospace projects as it can configure the key cost drivers and their corresponding attributes

Data Access Layer: An independent reusable layer can be built for database connections and interactions. It will be equipped with the required database technologies to accommodate an extensive dataset and integrate with different database systems in the near future.

STATE DIAGRAM:

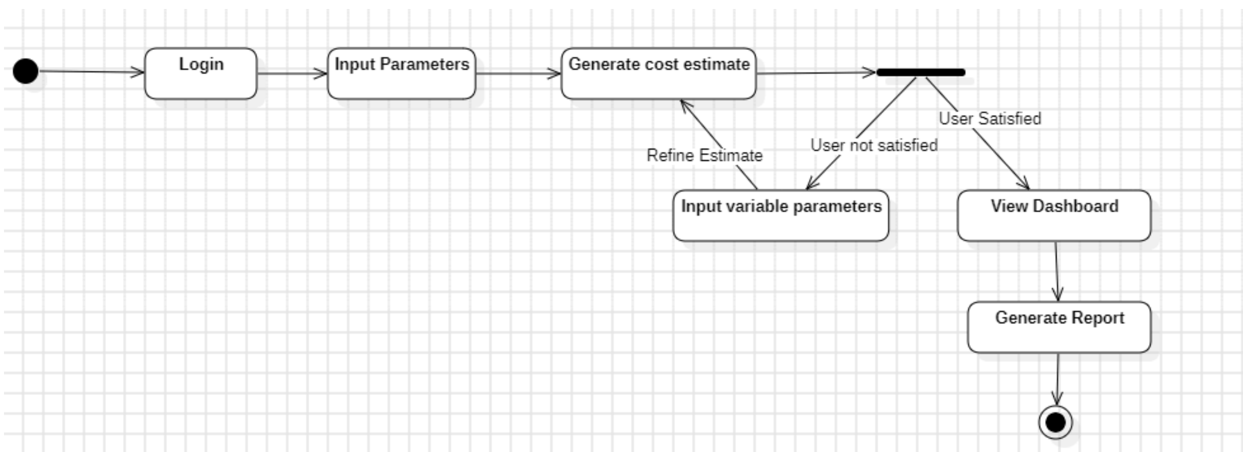


Figure 11. State Diagram

6.6 External Interfaces :

User Interface:

- The web-based interface allows aerospace engineers and project managers to interact efficiently with the system.
- Users can input the required project specifications, materials required, and other relevant manufacturing parameters through the interface.
- The dashboard displays cost estimation results, breakdowns, and top similar matches to users.

External Aerospace Manufacturing Data Sources:

- Integration with external aerospace manufacturing databases and systems to access industry-standard cost data, material prices, labor rates, and supply chain information.
- Data retrieval mechanisms to obtain relevant data for cost estimation, such as material prices, labor costs, and production lead times.

Authentication and Authorization Services:

- Authentication and authorization services ensure secure access control to the system.
- Users must authenticate themselves to access cost estimation functionalities and sensitive data.
- Permissions are granted based on user roles and responsibilities within the aerospace manufacturing organization.

Visualization Libraries:

- Utilization of visualization libraries for graphical representation of cost estimation results and analysis.

Feedback Collection Platforms:

- Implementation of a feedback system to gather user input and suggestions for system improvements.
- Existing feedback mechanisms or platforms may be integrated to collect feedback from aerospace engineers and stakeholders.

External Storage Services:

- Utilization of MongoDB or similar databases at the server level for data storage.

USE CASE DIAGRAM:

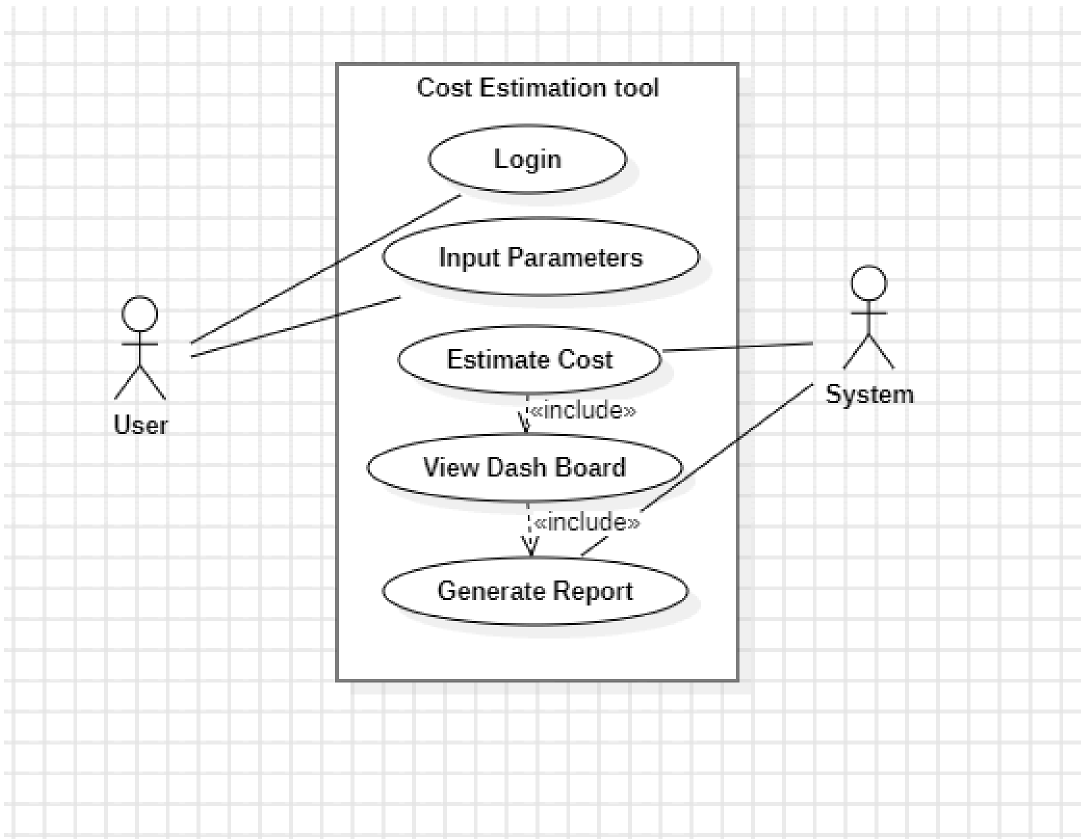


Figure 12. Use Case Diagram

6.7 Report Layouts

Report Description:

- The generated report provides a detailed analysis of cost estimation for aerospace manufacturing projects.
- It provides a comprehensive analysis to support informed decision-making.

Selection Criteria:

- Data included in the report is filtered based on project specifications, material requirements, and manufacturing parameters.
- Parameters such as material type, complexity, and production volume are considered in cost estimation.

Sorting and Grouping Criteria:

- The report data is primarily sorted by project identity to ensure clarity and traceability.
- Data is grouped based on analysis categories such as material required, labor costs, overhead expenses, and total project cost.
- Cost estimation results are further categorized by project phases or manufacturing stages for detailed analysis.

Tables Used:

- **Historical cost data of parts:** Provides insights into past cost trends and aids in comparing and analyzing current cost estimates
- **Part requirement sheets:** Contains detailed product specifications. It is essential for accurate cost estimation, material selection, and manufacturing process planning

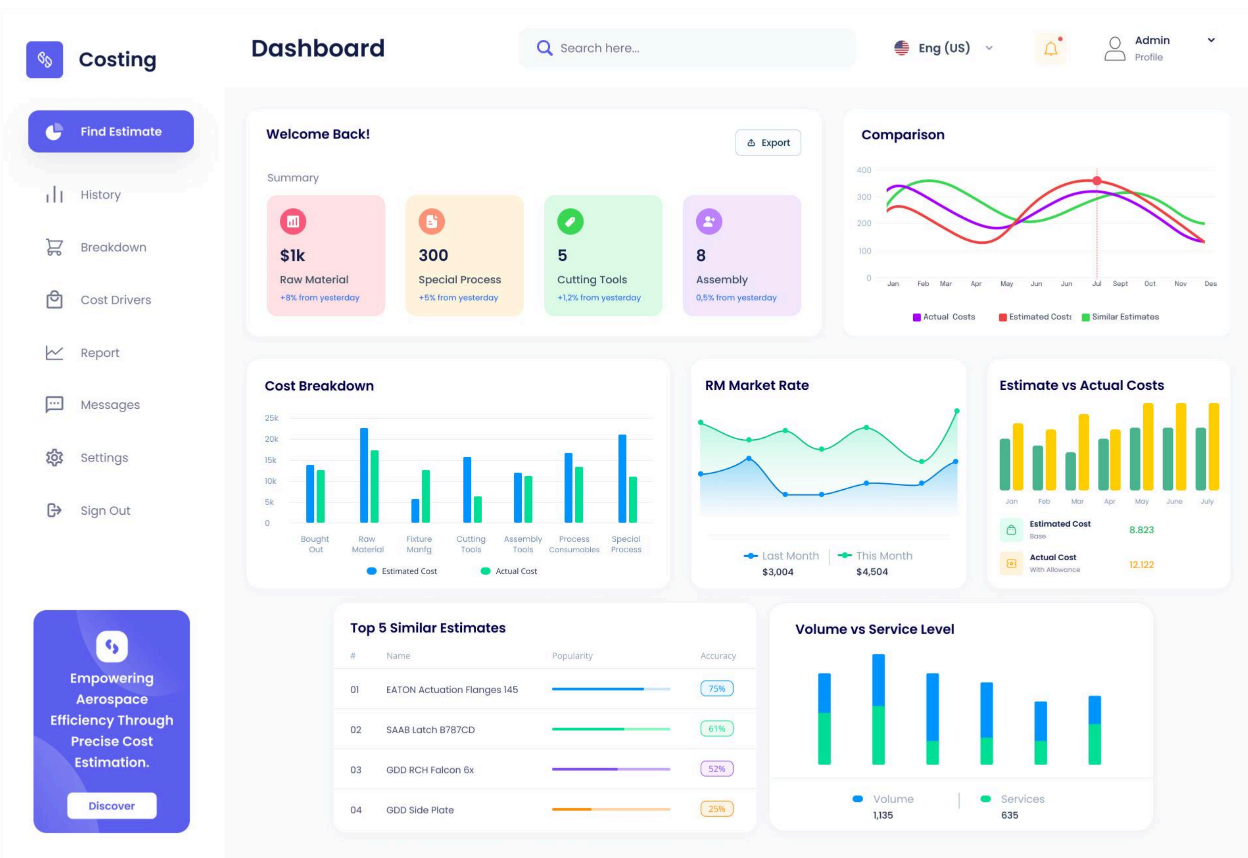


Figure 13. Dashboard Design

6.8 Deployment Diagram :

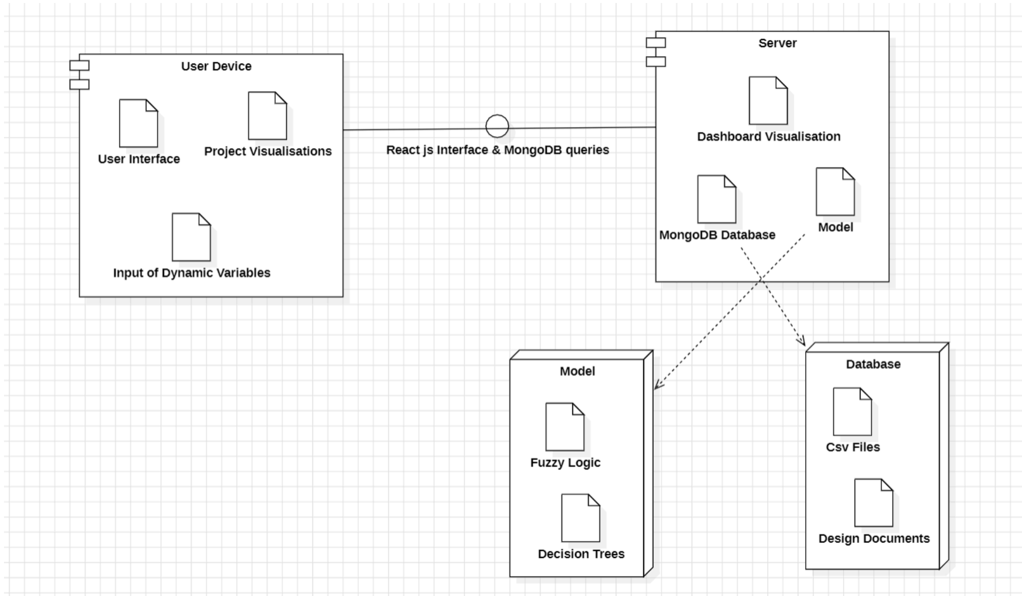


Figure 14. Deployment Diagram

6.9 Help:

- Online/Context-Sensitive Help: Integrated tooltips, pop-ups, or help buttons in the user interface.
- User Manual: Detailed guide on tool functionalities and workflows.
- Technical Manual: Documentation on tool architecture, algorithms, and data models.
- Knowledge Base: Articles, FAQs, and troubleshooting guides.

CHAPTER 7

IMPLEMENTATION AND PSEUDOCODE

- **<u>Clustering + Regression</u>:** Categorical attributes are first grouped and then within each group, we perform K-Means Clustering

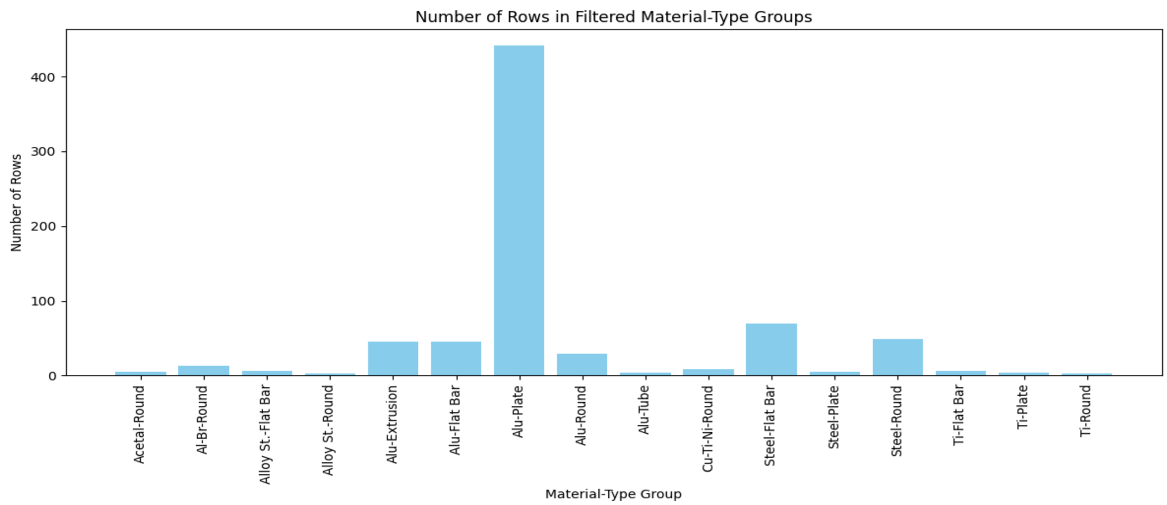


Figure 15. Grouping Method 1

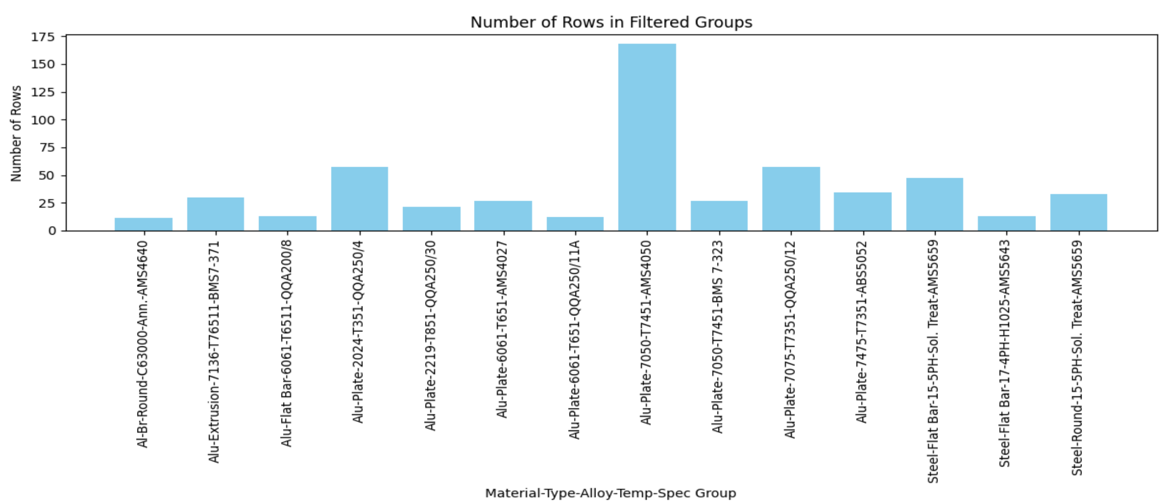
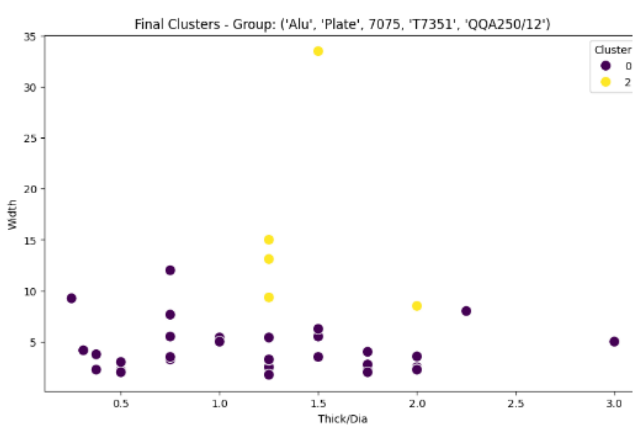


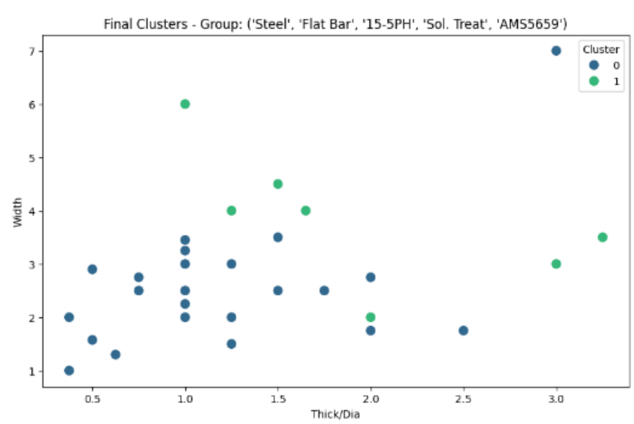
Figure 16. Grouping Method 2

```
Group: ('Al-Br', 'Round', 'C63000', 'Ann.', 'AMS4640')
Cluster Sizes:
Cluster 1: 9 data points
Cluster 2: 1 data points
Cluster 3: 1 data points
Merged cluster 2 with cluster 1
Merged cluster 3 with cluster 1

Cluster 1:
     Sl. # Material   Type   Alloy  Temp     Spec Thick/Dia Width  Length  \
194    197    Al-Br  Round  C63000  Ann.  AMS4640      0.75     0  24.000
252    255    Al-Br  Round  C63000  Ann.  AMS4640         1     0  24.000
265    268    Al-Br  Round  C63000  Ann.  AMS4640      1.25     0  24.000
311    315    Al-Br  Round  C63000  Ann.  AMS4640      1.75     0  24.000
318    322    Al-Br  Round  C63000  Ann.  AMS4640      1.75     0  24.000
463    471    Al-Br  Round  C63000  Ann.  AMS4640      0.75     0   0.818
688    828    Al-Br  Round  C63000  Ann.  AMS4640       1.5     0  24.000
691    831    Al-Br  Round  C63000  Ann.  AMS4640      4.75     0  24.000
703    845    Al-Br  Round  C63000  Ann.  AMS4640      1.25     0  24.000
704    846    Al-Br  Round  C63000  Ann.  AMS4640      1.25     0  24.000
705    847    Al-Br  Round  C63000  Ann.  AMS4640      1.25     0  24.000

     Weight in Kgs     RM Price  Price/Kilo  Cluster
194       1.318366    59.000000   44.752360        0
252       2.343762    64.580000   27.553990        0
265       3.662128    86.990000   23.753946        0
311       7.177772   193.000000   26.888568        0
318       7.177772   193.000000   26.888568        0
463       0.044934     0.231965    5.162323        0
688       5.273465   188.400000   35.726037        0
691      52.881134  4224.670000   79.889929        0
703       3.662128   109.460000   29.889722        0
704       3.662128   106.860000   29.179753        0
705       3.662128   113.500000   30.992906        0
Group: ('Alu', 'Extrusion', 7136, 'T76511', 'BMS7-371')
Cluster Sizes:
Cluster 1: 17 data points
Cluster 2: 1 data points
Cluster 3: 12 data points
```

Figure 17. Screenshot of clusters formed

Few of the Results:





Clustering Results For All Groups

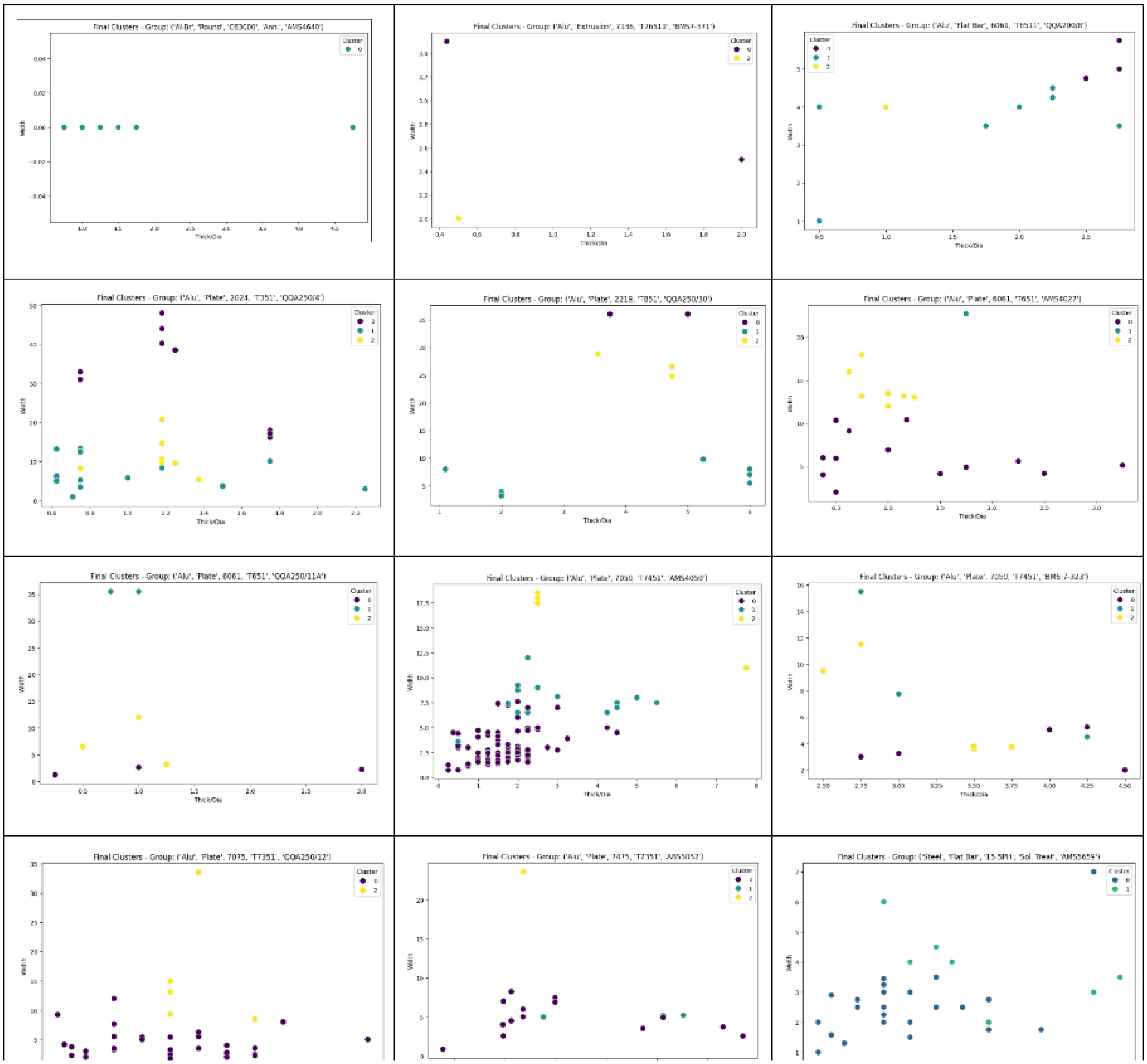


Figure 18. Visualizations of the clusters formed

Pseudocode:

```
from sklearn.cluster import KMeans
import numpy as np

# Define the number of clusters
n_clusters = 3

# Iterate over the filtered groups
for group_name, group_data in grouped:
    if group_name in filtered_groups:
        print(f"Group: {group_name}")

        # Select relevant features for clustering
        features = group_data[['Thick/Dia', 'Width', 'Length', 'Weight in Kgs']]

        # Perform K-means Clustering
        kmeans = KMeans(n_clusters=n_clusters, random_state=42)
        clusters = kmeans.fit_predict(features)

        # Add cluster labels to the data
        group_data['Cluster'] = clusters

        # Print the number of data points in each cluster
        cluster_counts = group_data.groupby('Cluster').size()
        print("Cluster Sizes:")
        for cluster, count in cluster_counts.items():
            print(f"Cluster {cluster + 1}: {count} data points")

        # Print the clusters within the group
        unique_clusters = np.unique(clusters)
        for cluster in unique_clusters:
            cluster_data = group_data[group_data['Cluster'] == cluster]
            print(f"\nCluster {cluster + 1}:")
            print(cluster_data)
```

- **<u>Fuzzy logic</u>:** Numerical attributes are initially created into linguistic variables using quartiles by categorizing them

```
import pandas as pd

# Read the data into a DataFrame
df = pd.read_csv('/content/RMcostdetails.csv.csv')

# Convert 'Length' column to numeric
df['Length'] = pd.to_numeric(df['Length'], errors='coerce')

# Convert 'Width' column to numeric
df['Width'] = pd.to_numeric(df['Width'], errors='coerce')

# Convert 'Weight in Kgs' column to numeric
df['Weight in Kgs'] = pd.to_numeric(df['Weight in Kgs'], errors='coerce')

# Convert 'Thick/Dia' column to numeric
df['Thick/Dia'] = pd.to_numeric(df['Thick/Dia'], errors='coerce')

# Define thresholds for categorization
high_threshold = df['Thick/Dia'].quantile(0.7)  # You can adjust the quantile value based on your preference
medium_threshold = df['Thick/Dia'].quantile(0.3)  # You can adjust the quantile value based on your preference

# Define a function to categorize Thick/Dia values
def categorize_thick_dia(thick_dia):
    if thick_dia >= high_threshold:
        return 'High'
    elif thick_dia >= medium_threshold:
        return 'Medium'
    else:
        return 'Low'

# Apply the categorization function to the Thick/Dia column
df['Thick/Dia_Category'] = df['Thick/Dia'].apply(categorize_thick_dia)

# Define thresholds for categorization
high_threshold = df['Width'].quantile(0.7)  # You can adjust the quantile value based on your preference
medium_threshold = df['Width'].quantile(0.3)  # You can adjust the quantile value based on your preference

# Define a function to categorize width values
def categorize_width(width):
    if width >= high_threshold:
        return 'High'
    elif width >= medium_threshold:
        return 'Medium'
    else:
        return 'Low'

# Apply the categorization function to the width column
df['Width_Category'] = df['Width'].apply(categorize_width)
```

Creation of membership functions

```
import numpy as np
import skfuzzy as fuzz
from skfuzzy import control as ctrl

# Define the linguistic variables
thickness = ctrl.Antecedent(np.arange(df['Thick/Dia'].min(), df['Thick/Dia'].max() + 1, 1), 'thickness')
width = ctrl.Antecedent(np.arange(df['Width'].min(), df['Width'].max() + 1, 1), 'width')
length = ctrl.Antecedent(np.arange(df['Length'].min(), df['Length'].max() + 1, 1), 'length')
weight = ctrl.Antecedent(np.arange(df['Weight in Kgs'].min(), df['Weight in Kgs'].max() + 1, 1), 'weight')

# Define membership functions for thickness
thickness['Low'] = fuzz.trimf(thickness.universe, [df['Thick/Dia'].min(), df['Thick/Dia'].min(), medium_threshold])
thickness['Medium'] = fuzz.trimf(thickness.universe, [df['Thick/Dia'].min(), medium_threshold, high_threshold])
thickness['High'] = fuzz.trimf(thickness.universe, [medium_threshold, high_threshold, df['Thick/Dia'].max()])

# Define membership functions for width
width['Low'] = fuzz.trimf(width.universe, [df['Width'].min(), df['Width'].min(), medium_threshold])
width['Medium'] = fuzz.trimf(width.universe, [df['Width'].min(), medium_threshold, high_threshold])
width['High'] = fuzz.trimf(width.universe, [medium_threshold, high_threshold, df['Width'].max()])

# Define membership functions for length
length['Low'] = fuzz.trimf(length.universe, [df['Length'].min(), df['Length'].min(), medium_threshold])
length['Medium'] = fuzz.trimf(length.universe, [df['Length'].min(), medium_threshold, high_threshold])
length['High'] = fuzz.trimf(length.universe, [medium_threshold, high_threshold, df['Length'].max()])

# Define membership functions for weight
weight['Low'] = fuzz.trimf(weight.universe, [df['Weight in Kgs'].min(), df['Weight in Kgs'].min(), medium_threshold])
weight['Medium'] = fuzz.trimf(weight.universe, [df['Weight in Kgs'].min(), medium_threshold, high_threshold])
weight['High'] = fuzz.trimf(weight.universe, [medium_threshold, high_threshold, df['Weight in Kgs'].max()])

# Visualize the membership functions (optional)
thickness.view()
width.view()
length.view()
weight.view()
```

Linear Regression:

```
# Step 3: Convert columns to numeric, replacing non-numeric values with NaN
data['Thick/Dia'] = pd.to_numeric(data['Thick/Dia'], errors='coerce')
data['Width'] = pd.to_numeric(data['Width'], errors='coerce')
data['Length'] = pd.to_numeric(data['Length'], errors='coerce')
data['Weight in Kgs'] = pd.to_numeric(data['Weight in Kgs'], errors='coerce')

# Drop rows with NaN values
data.dropna(subset=['Thick/Dia', 'Width', 'Length', 'Weight in Kgs', 'RM Price'], inplace=True)

# Assuming 'RM Price' is the target variable, and other columns are features
X = data[['Thick/Dia', 'Width', 'Length', 'Weight in Kgs']]  # Features
y = data['RM Price']  # Target variable

# Step 4: Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Step 5: Train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Step 6: Evaluate the model
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print('Mean Squared Error:', mse)

# Step 7: Visualize the results
plt.scatter(y_test, y_pred)
plt.xlabel('Actual')
plt.ylabel('Predicted')
plt.title('Actual vs. Predicted')
plt.show()
```

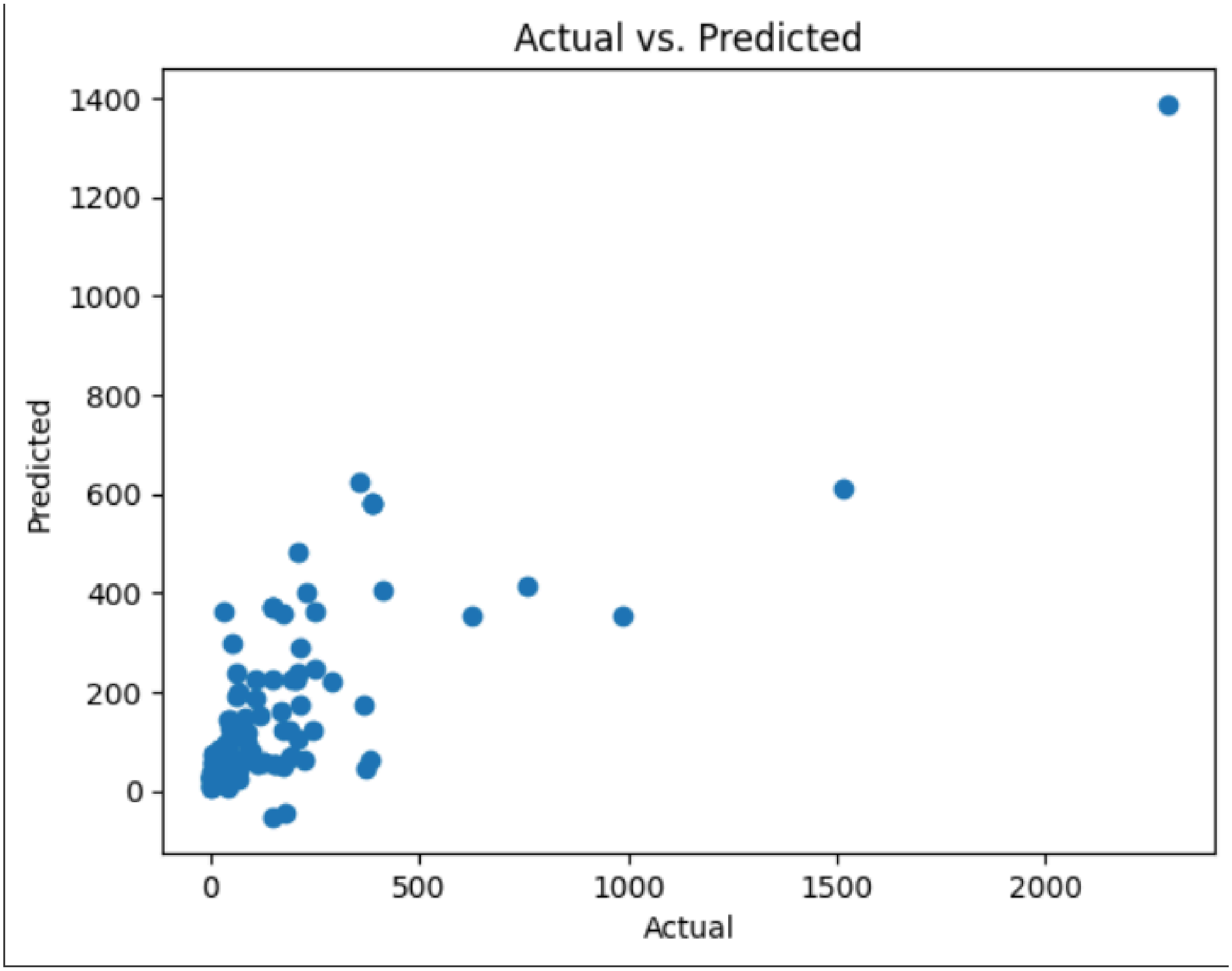


Figure 19. Relationship between predicted and actual values

```
from sklearn.metrics import r2_score

# Evaluate the model using R2 score
r2 = r2_score(y_test, y_pred)
print('R-squared (R2) Score:', r2)
```

```
R-squared (R2) Score: 0.6308005139222775
```

the linear regression model explains around 63.18% of the variance in the dependent variable (RM Price) based on the independent variables (Thick/Dia, Width, Length, and Weight in Kgs).

CHAPTER 8
CONCLUSION OF CAPSTONE PROJECT PHASE - 1

We have completed about 30% of our project so far

1) We conducted extensive research and analysis to determine the technologies and methodologies that we can use during the various phases of our project.
2) A good amount of progress has been made in preparing a comprehensive dataset for our project. Some of the features our dataset comprises of are Product,Part Image,Raw Material size, Alloy, Number of parts,Set up time and many more features
3) We have also successfully completed the preliminary requirements specification (PRS) document and high-level design (HLD) document, which play an important role in the overall project process.
4) We also continue to gain an understanding of the aerospace manufacturing industry and the processes involved in manufacturing of a component.
5) We have also been meeting with industry experts to gain more information and increase our domain knowledge about the aerospace manufacturing industry

CHAPTER 9

PLAN OF WORK FOR CAPSTONE PROJECT PHASE - 2

With the completion of the documentation as well as the design phases as well as the basic collection of the dataset for the application, we will have successfully completed phase 1. During phase 2 of our project, We will mainly be focused on the implementation of the model, testing out which approach works the best, refining based on the feedback received, and finally submitting a paper. We plan on doing so by dividing our tasks as follows-

- We will start out by selecting the most appropriate machine learning techniques needed to build the cost estimation model
- We will then implement these models and determine which would be the best one to use among the several machine learning techniques which we thought would work
- We will also look into further refining the features in our dataset to gain better predictive analytics
- The splitting of our dataset into training and validation sets will also be carried out
- We wish to create our user interface and enhance overall design of the application based on feedback received. This would involve making the user experience optimized and much more pleasing to the eye of our user
- The implementation of visualizations showcasing trends related to cost estimations will also be looked into

APPENDIX A: DEFINITIONS, ACRONYMS AND ABBREVIATIONS

DEFINITIONS:

1. **Cost Estimation Model:** A computational model designed to predict the manufacturing costs of aerospace components based on input parameters such as material type, dimensions, complexity, and manufacturing process.
2. **User Interface (UI):** The graphical interface through which users interact with the cost estimation system, providing input parameters and receiving cost estimation results.
3. **Deep Learning:** A subset of machine learning algorithms that use artificial neural networks with multiple layers to extract high-level features from data and make predictions.
4. **Fuzzy Decision Trees:** Decision trees augmented with fuzzy logic to handle uncertainty and imprecision in input data for more accurate cost estimations.
5. **Defuzzification:** The process of converting fuzzy estimations generated by fuzzy decision trees into precise numerical values for clearer interpretation and decision-making.

ABBREVIATIONS:

1. **UI:** User Interface
2. **CNNs:** Convolutional Neural Networks
3. **APIs:** Application Programming Interfaces
4. **XML**: Extensible Markup Language
5. **GPU:** Graphics Processing Unit
6. **PRS:** Part Requirement Sheet
7. **HTTP:** Hypertext Transfer Protocol
8. **FAQs:** Frequently Asked Questions

REFERENCES

[1] Jung JY. Manufacturing cost estimation for machined parts based on manufacturing features. Journal of intelligent manufacturing. 2002 Aug;13:227-38.

[2] Psarommatis F, Danishvar M, Mousavi A, Kiritsis D. Cost-based decision support system: a dynamic cost estimation of key performance indicators in manufacturing. IEEE Transactions on Engineering Management. 2022 Jan 7.

[3] Papatheocharous E, Andreou AS. A hybrid software cost estimation approach utilizing decision trees and fuzzy logic. International Journal of Software Engineering and Knowledge Engineering. 2012 May;22(03):435-65.

[4] Ji SH, Ahn J, Lee HS, Han K. Cost estimation model using modified parameters for construction projects. Advances in Civil Engineering. 2019 Jul 30;2019.

[5] Ning F, Qu H, Shi Y, Cai M, Xu W. Feature-Based and Process-Based Manufacturing Cost Estimation. Machines. 2022 Apr 28;10(5):319.

[6] Bao HP, Samareh JA, Weston RP. Predicting production costs for advanced aerospace vehicles. In61st Annual Conference of Society of Allied Weight Engineers, Inc. 2002 Jan 1 (No. SAWE Paper-3246).

[7] Bacharoudis K, Wilson H, Goodfellow-Jones S, Popov A, Ratchev S. An efficient cost estimation framework for aerospace applications using matlab/simulink. Procedia CIRP. 2021 Jan 1;104:1143-8.

[8] Díaz A, Fernández S, Guerra L, Díaz E. Manufacturing cost prediction through data mining. InDevelopments and Advances in Defense and Security: Proceedings of MICRADS 2020 2020 (pp. 251-258). Springer Singapore.

[9] Chansaad AP, Chaiprapat S, Yenradee P. A parametric-fuzzy logic system for paint cost estimation of products with uncertain geometric characteristics. In2013 International Conference on Fuzzy Theory and Its Applications (iFUZZY) 2013 Dec 6 (pp. 482-487). IEEE.

[10] Papatheocharous E, Andreou AS. A hybrid software cost estimation approach utilizing decision trees and fuzzy logic. International Journal of Software Engineering and Knowledge Engineering. 2012 May;22(03):435-65.

[11] Baioumy, S. Product Cost Estimation Using Fuzzy Logic. The International Conference on Applied Mechanics and Mechanical Engineering, 2010; 14(14th International Conference on Applied Mechanics and Mechanical Engineering.): 1-12. doi: 10.21608/amme.2010.37648