**DETAILED PROJECT REPORTS**

**THE WAFER FAULT DETECTION PROJECT**

**CONTENTS**

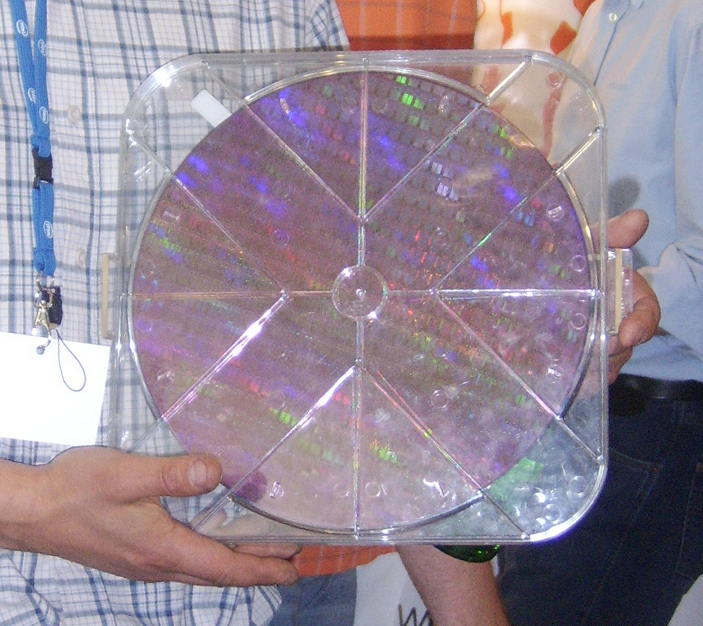
1. Introduction
2. Problem Statement and Data Description
3. The Application Flow

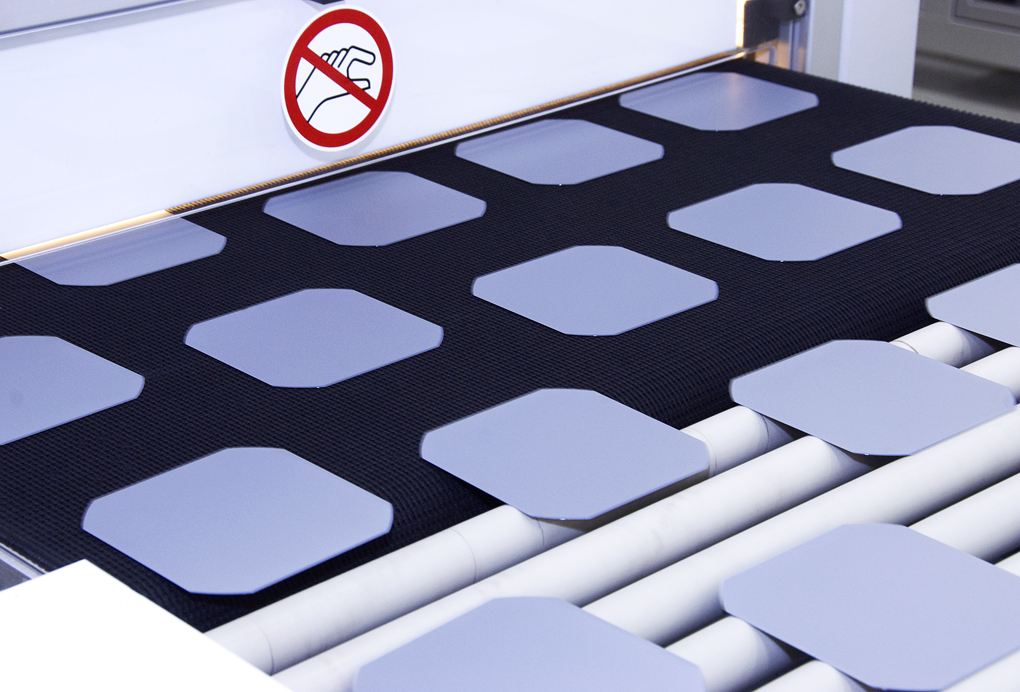
* Code and Logging
* Validation and Transformation
* DB Operations
* Data Processing
* Clustering
* Model Selection and Tuning
* Prediction
* Deployment

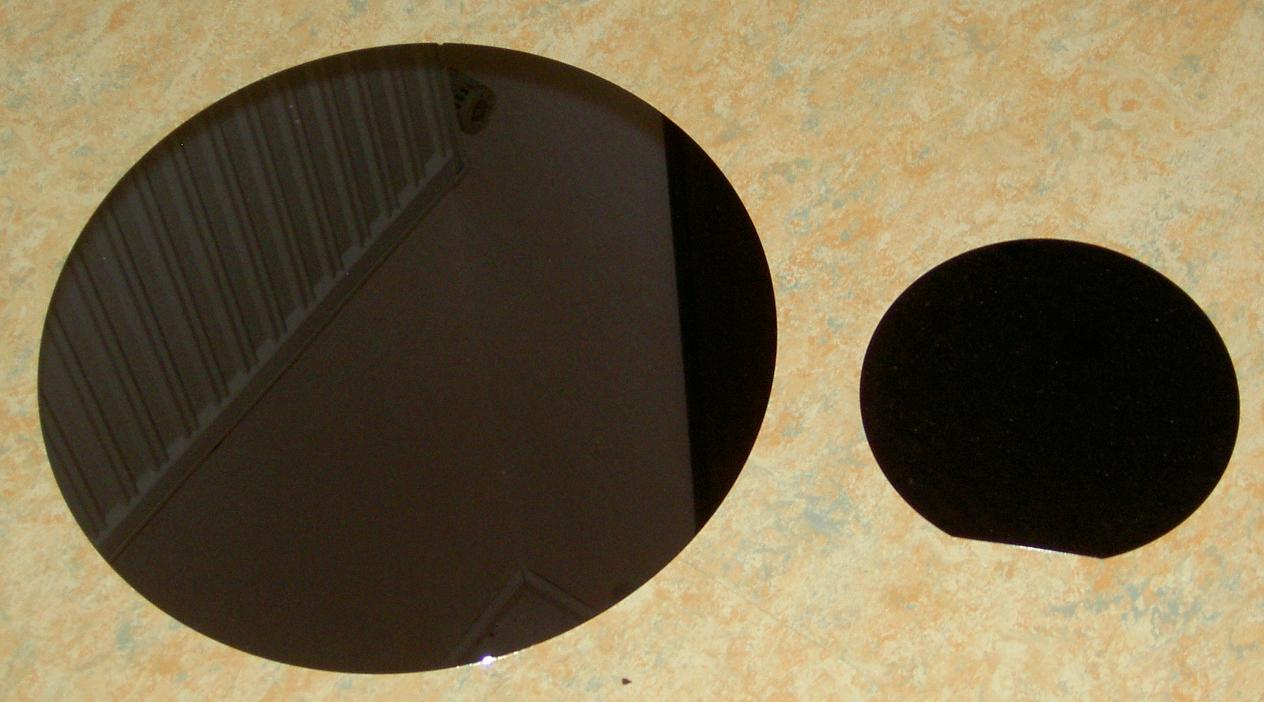
**1.Introduction**

**Wafer**

In [electronics](https://en.wikipedia.org/wiki/Electronics), a wafer (also called a slice or substrate) is a thin slice of [semiconductor](https://en.wikipedia.org/wiki/Semiconductor), such as a [crystalline silicon](https://en.wikipedia.org/wiki/Crystalline_silicon) (c-Si), used for [the fabrication](https://en.wikipedia.org/wiki/Semiconductor_device_fabrication) of [integrated circuits](https://en.wikipedia.org/wiki/Integrated_circuit) and, in [photovoltaics](https://en.wikipedia.org/wiki/Photovoltaics), to manufacture [solar cells](https://en.wikipedia.org/wiki/Solar_cell). The wafer serves as the [substrate](https://en.wikipedia.org/wiki/Substrate_(materials_science)) for [microelectronic](https://en.wikipedia.org/wiki/Microelectronic) devices built in and upon the wafer. It undergoes many [microfabrication](https://en.wikipedia.org/wiki/Microfabrication) processes, such as [doping](https://en.wikipedia.org/wiki/Doping_(semiconductor)), [ion implantation](https://en.wikipedia.org/wiki/Ion_implantation), [etching](https://en.wikipedia.org/wiki/Etching_(microfabrication)), [thin-film deposition](https://en.wikipedia.org/wiki/Thin-film_deposition) of various materials, and [photolithographic](https://en.wikipedia.org/wiki/Photolithography) patterning. Finally, the individual microcircuits are separated by [wafer dicing](https://en.wikipedia.org/wiki/Wafer_dicing) and [packaged](https://en.wikipedia.org/wiki/Integrated_circuit_packaging) as an integrated circuit.

 fig.1.1

 fig.1.2

 fig.1.3

A wafer is very thin slice of semiconductor material which is used in various electronics circuit and devices.

**2. Problem Statement and Data Description**

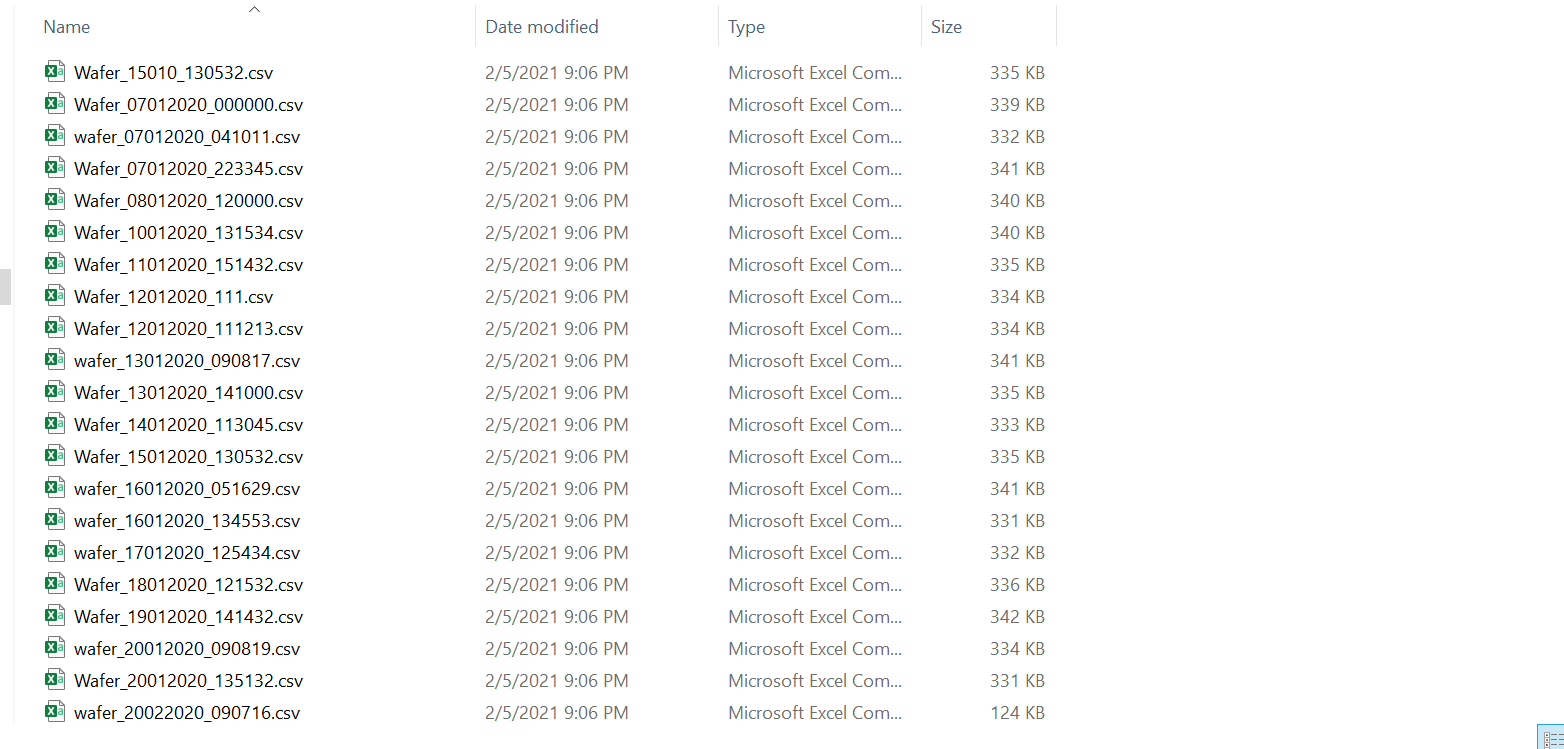
**Problem Statement**

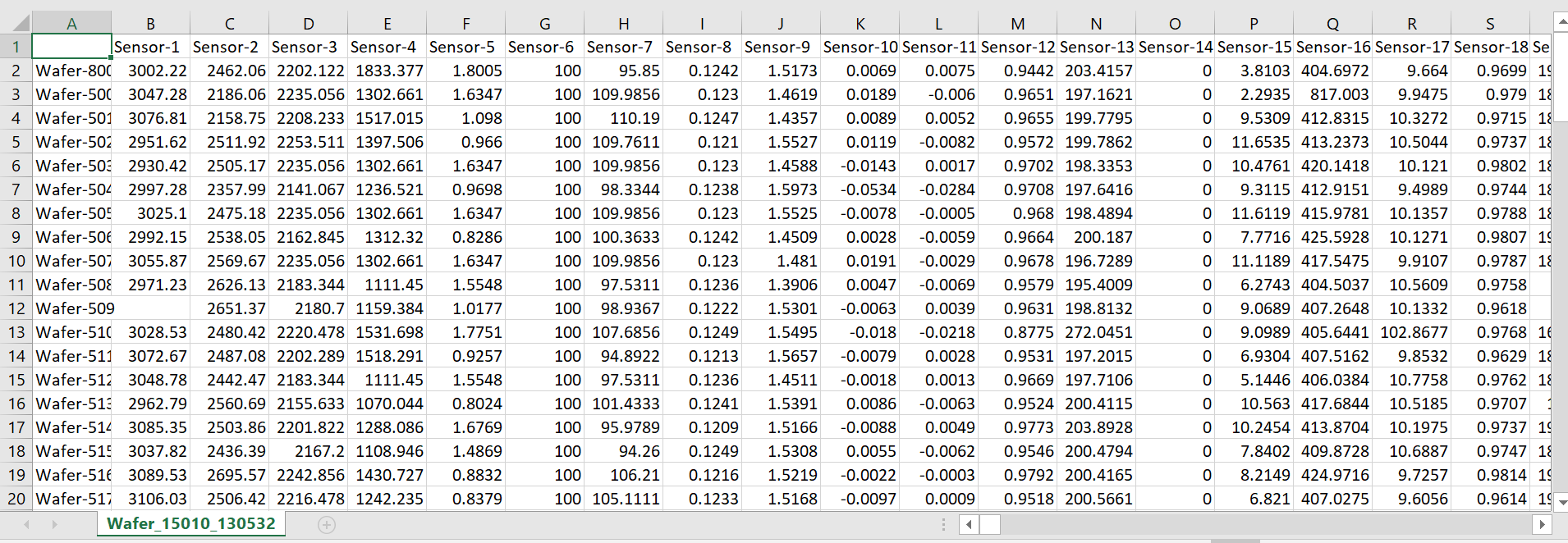
Our client has production line, and the production line has various no wafer are deployed in this whenever any fault occurs anywhere in the any wafer then they have to stope entire pipeline and check it manually and then they replace it so whole production line got stoked.

Our client has asked us to create a model which successfully identify that which section of pipeline has faulty wafer so only that part of pipeline will be closed and whole process will be continue without any hindrance and that faulty wafer will get replace and whole process of production pipeline will continue again.

**“To build a classification methodology to predict the quality of wafer sensors based on the given training data.”**

**Data Description**

fig 2.1

 fig 2.2

There was nineteen csv file each file has almost 120 record and feature no is “B” to “VS” that means there is 90 features.

Formats of data file “Wafer\_DDMMYYYY\_HHMMSS.CSV”

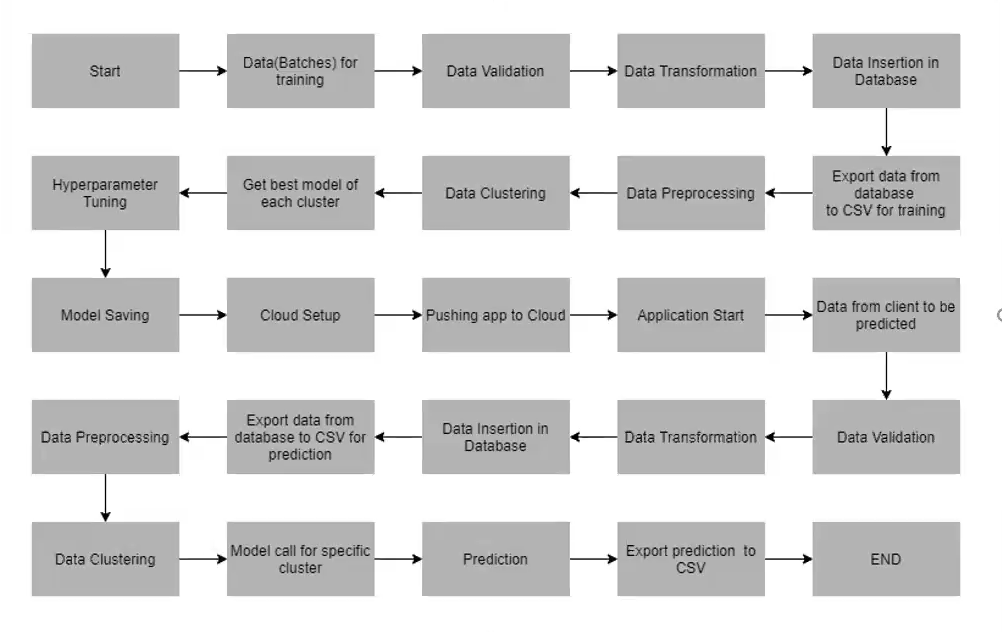
**3. The Application Flow**

Application flow shows the steps to be follow for end result. As we can see so many steps are defined in the given flow diagram (fig.3.1), which we implemented in our model creation.

To make our model we divided our work in different-different steps to complete Reason for that each work will get easy if any error comes in any module, we can modify easily, or any modification needed we modify easily or it possible we work as a team work will get divided and work can do smoothly.

Steps from Data (Batches) for training to Data Insertion in Database is called **data Injection Pipeline,** in data(batches) for database this is the location of dataset where the data is stored to train it is provided by client. Data Validation it is process of validation according to our **DSA (**data sharing agreements) that, it is agreement between client and model creation contract that with term and responsibilities. Format of dataset, timing, frequency of data etc.

Data Transformation is the process of converting data from one format to another, typically from the format of a source system into the required format of a destination system. Data transformation is a component of most data integration and data management tasks, such as data wrangling and data warehousing. Data Insertion in Database after data validation and transformation we create two files where we store our data one is wrong data file and other one is useful one and we use further useful file to train our model and with wrong file we discuss with client for issue and other application.

fig.3.1

Steps from Export data from database to csv for training to Model saving can be known as **model training**, Export data from database to csv for training is basically select the useful files of dataset this work as input of training the model, Data Preprocessing is performing different work like is there any null dataset, distribution of dataset is normal or not, handle categorical dataset etc. Data Clustering after preprocessing we apply clustering algorithm to make clusters based on feature similarity here, we just don’t want make single model we want make one than one model so that our efficiency of model get increase. Get best Model of each cluster after applying clustering we make separate models for each clustering so can we can get better performance of model. Hyperparameter Tuning it is process of deciding factor inside model used algorithms so we can get more efficiency. After applying hyperparameter tuning we save the model for further work that is Model saving.

Steps from Cloud setup to Application Start is can be called **Model Deployment** there is many platforms available for deployment of model like AWS, GCP, Azure, pivotal etc. for development we will need an web application to work with like flask Api, FAST Api etc. Every platform has its own pros and cons and modification needed in code file. Pushing the code at cloud platform is just putting the code of model on cloud and after that we can start out model.

Now our model is ready for prediction but we cannot put file directly dataset for prediction so we need to preprocess all testing/prediction dataset as training dataset we processed. Again, after we predict the output.

Main.py

1.Validation 2. Training

Read Data Read Training Data

Validation Data preprocessing

Transformation Clustering

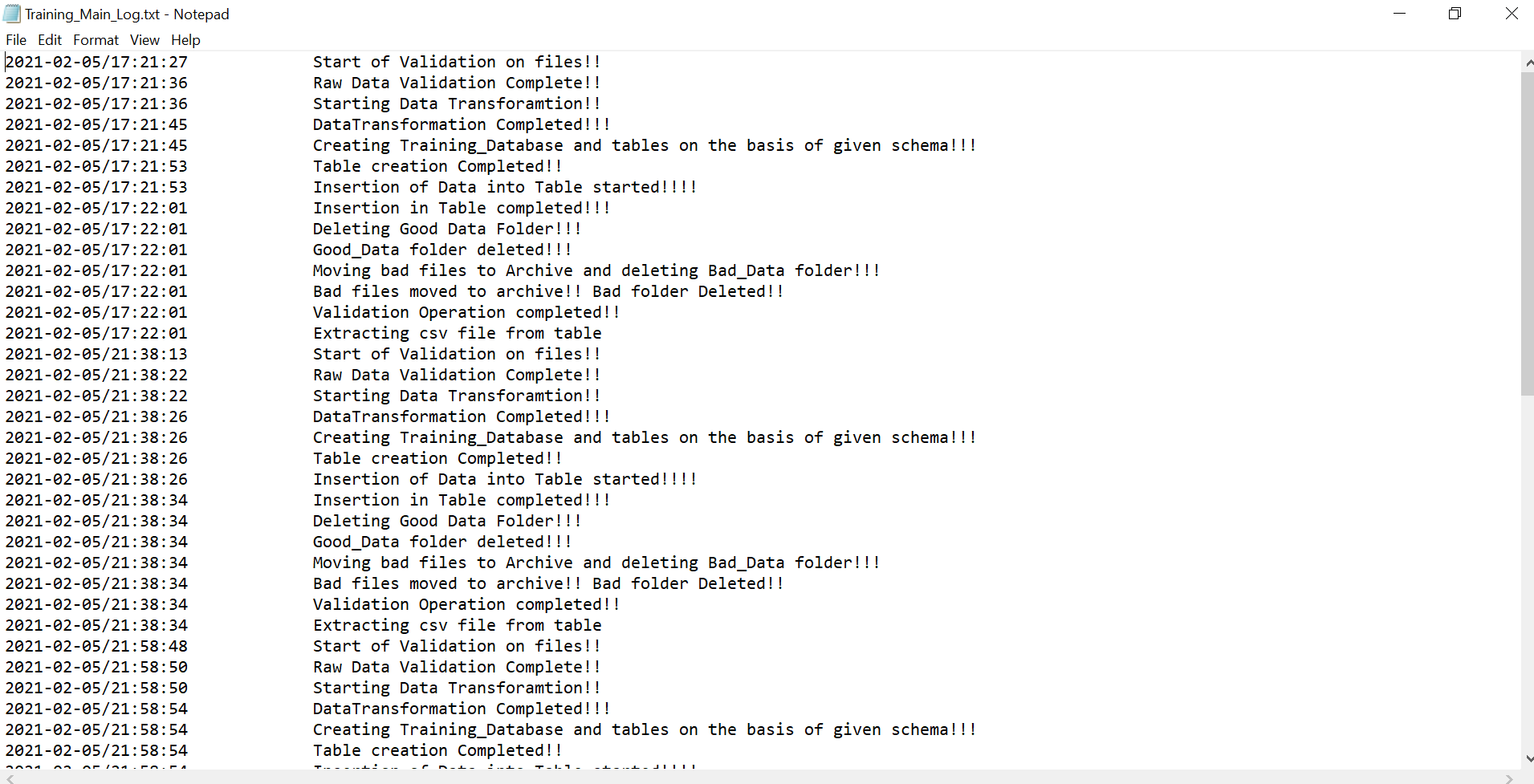
Insertion Model creating

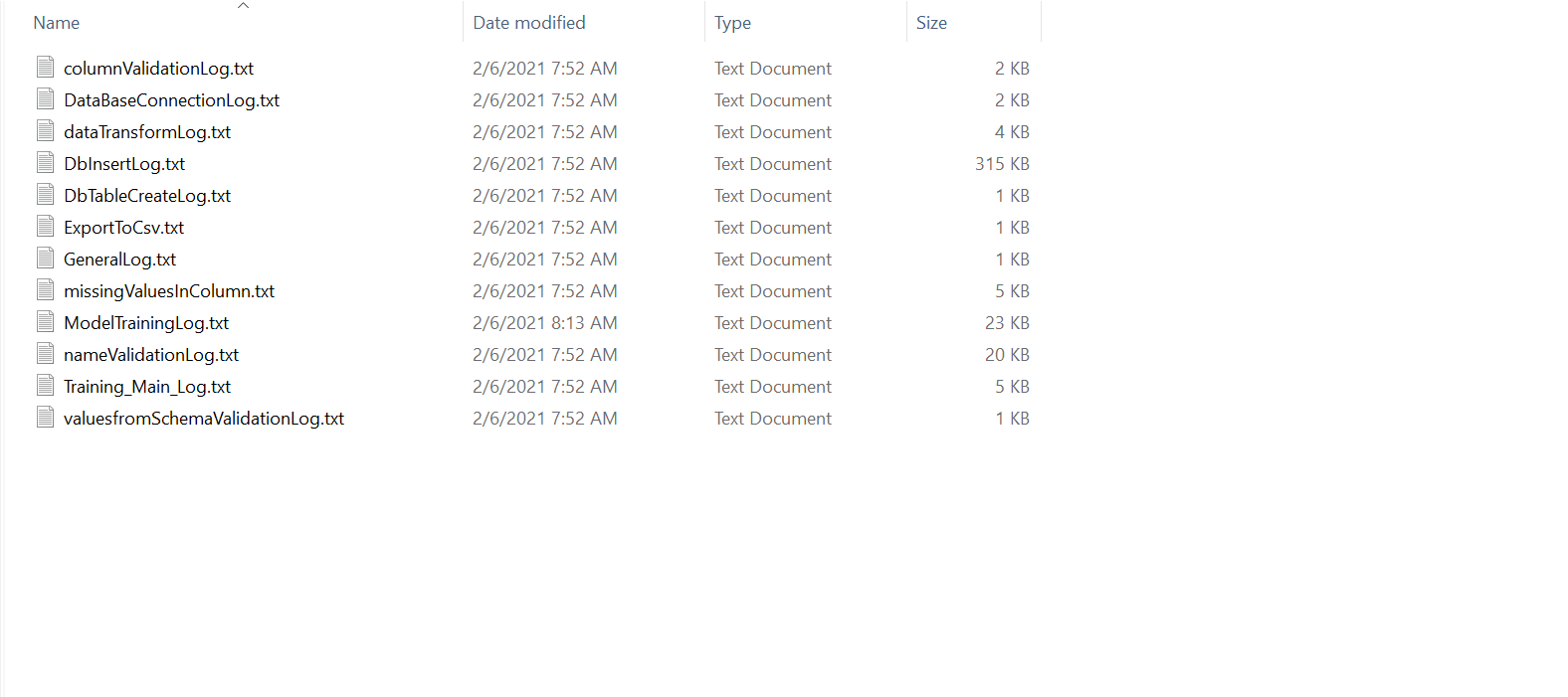
Export csv Model tuning

Deploy

**Code and Logging**

This is basically coding part that handle the process statement which give message according to logging that what happened in that particular step. Like file insertion successful, stored in wrong file dataset if formats are no correct basically it is the step by which we can monitor on activities of model. Without much trouble any if any error happens, we can easily get to know what error are coming from which section is it with code, model or dataset.

fig.4.1

 fig.4.2