

# Detail Project Report

Project Name : **Sensor fault detection**

By:

Mukesh Kumar Sharma

+91 8106940790

## Content:

### 1. Project Description

- a. Time taken to build the model
- b. Team members

### 2. Data Part

- a. Sample data preparation

### 3. Development strategies

- a. Database
- b. Mlflow
- c. Data Version Control (DVC)

### 4. Productization

- a. API
- b. Cloud

### 5. Test case

- a. Local machine

### 6. Technical Requirements

## **Project Description:**

Project Title: Sensor Fault Detection

It is an industry based project which is used for detecting the fault sensor in vehicle. As we know that in today's generation every new model or version of vehicle have been developed by the engineer and for the proper functionality of the entire vehicle they have been using sensor's.

In this project I have collected multiple series of sensor information and the main idea behind developing this model is to automate the operation of any vehicle and also help in service center, whenever any person comes to the service center for vehicle service, the technician need to check each and every sensor and it's a time consumption process. To overcome this problem I have imposed a machine learning model, where by just imposing this model it will automatically detect the fault sensor and technician will easily change the sensor.

The main idea behind developing this model is to reduce the effort of the technician and save lot's of time.



Time Taken : To develop this model I along with my team It has taken two months as it was our first model and we have faced many challenges such as

- Information collection related to the project
- Data gathering, merging, etc.
- Execution planning
  - UI/UX planning
  - Deployment planning
  - Model development
  - Testing

Each and every task was divided into the team members and the time limit was set. In this way It took two months to completion of this project.

Time Distribution:

- Information collection related to the project → 1 week
- Data gathering, merging, etc. → 1 week
- Execution planning → 2 week
- Model development and deployment → 2 week
- Testing → 2 week

Total 10 people were involved in developing this model and my role was to **clean the data, develop multiple model, checking the development in local machine.**

## Data Part :

In this Data part we have randomly collected few data from the bunch of data and consider this as a testing data. **In this stage my role is to check whether the model is working with good accuracy or not using this testing data.**

Once this is done then only we will be going to deploy the model in cloud machine.

## Development strategies :

To develop the model we are going to use the following :

**Databse** that we are going to use in MongoDB and SQL

**CICD pipeline** which is a method that builds, tests and deploys code by using automation and also help to automates your software delivery process.

**DevOps** is also used in the model for efficiently compile, built and deploy their code to their production compute platform.

**MLFlow** is also involved in this model development as to handle and manage the end to end machine learning life cycle.

*Note: Each and every line of code is written in Modular Fashion which will improve the reusability of code.*

## Productization :

To productize this model we have first test the model in the local machine using **Flask API** where the input is passed from the local user and checking the output and monitoring the accuracy.

**Once we have tested in the local machine then we have deployed in the cloud using GCP, AWS, and AZURE**

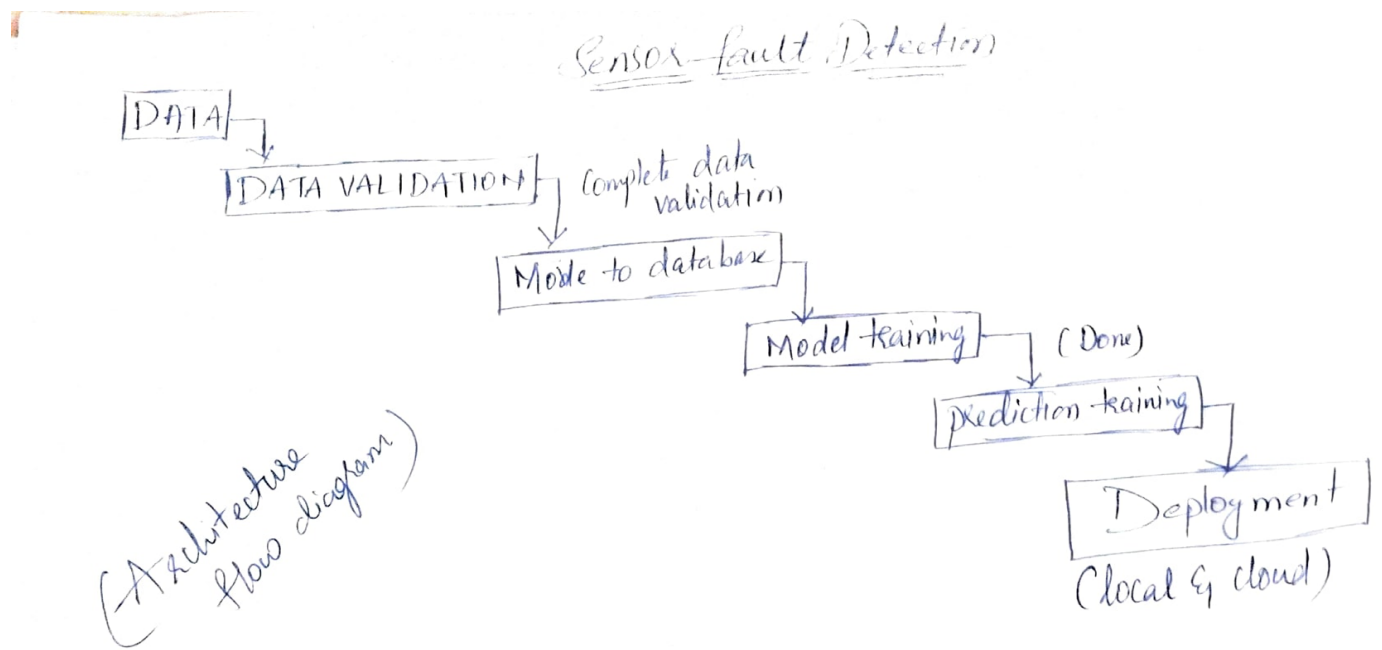
Post Productization :

In this state we have retrained the model as if their was a small bug or loop hole where there was a chance of improper train of model. So we have retrain the model.

Technical aspect:

1. ML algorithm
2. CICD with DVC
3. MLFlow
4. Flask
5. Cloud deployment

## Architecture design :



## FAQ

Q1. Tell me about your current project.

Ans: The project is called ***Sensor Fault Detection*** using Machine Learning.

The goal is to identify the fault sensor in a vehicle, as we know that in today's generation any new model of vehicle launched into the market are coming with lot's of sensors. Let's say it as in car for parking we will have different sensor, then for dashboarding access we will have sensors, say breaking, turning, reverse, gear system each and every components in the vehicle consists of sensors. Imagine that any one of the sensor fault all of a sudden then we need to look at each of the sensors, then check them one by one after that if any error occur find it and directly replace it.

For model to work properly I have first used ***k-means clustering for grouping similar sensors and then used xgboost for prediction.***

Q2. What was the size of the data?

Ans: The size of my training data is a multiple series of sensors batch file in .csv format ***and its length is 590.***

Q3. What was the data type?

Ans: My datatype consists of ***one character column, sensors column = 590 which is of float datatype and one Output column that hold integer datatype.***



Q4. What was the team size and distribution?

Ans: My entire team consists of

- UI/UX developer for developing the front end.
- Deployment eng. for deploying the model
- Model developer
  - Data scientist → modeling purpose
  - Coder → python developer
  - Data base handler → MongoDB
- Testing
  - Tool tester
  - Unit testing
- Mentor to allocate task and validate it

Q5. How were you creating and maintaining the logs?

Ans: All the logs are captured at each step of execution and stored in Data base directly.

Q6. What techniques were you using for data pre-processing?

Ans: 1. check all the sensor batch file whether they are in specific format or not, if yes then move to good data folder. 2. Handling missing value using KNNImputer. 3. Also create a manual regex function and stored the actual structure of the data for automation purpose.

Q7. How were you maintaining the failure cases?

Ans: All the failure data will be moved to bad data folder and we will not be using this for prediction purpose.

Q8.What kind of automation have you done for data processing?

Ans: I have created a class function and stored all the required operation that are need in preprocessing say ***checking and handling missing value, removing unwanted columns, data seperation.***

Q9.How are you monitoring your job?

Ans: For monitering purpose I have used logging function which will automatically log each and every operation at the time of execution.

Q10. What were your roles and responsibilities in the project?

Ans: My roles and responsibilities was the following:

- a. Creating a preprocessing function that can easily preform ***checking and handling missing value, removing unwanted columns, data seperation***
- b. Developing the model using ***xgboost algorithm.***

Q11.What was your day to day task?

Aas: My day to day task was to coordinate with the team members, conduction meeting, code distribution, handling github, and model development.

Q12.In which area you have contributed the most?

Ans: I have mostly worked in preprocessing stage and simple model building using xgboost.

Q13. In how many projects you have already worked?

Ans: I have worked on multiple project say Insurance price prediction, credit card detection, chatbot, face matching using Deep learning (basic) and wine quality.

Q14. How were you doing deployment?

Ans: Deployment is completely based upon mentor requirement.

Q15. What kind of challenges have you faced during the project?

Ans: The major challenges that I have faced in personal is to deal with series of data batches as there are multiple batch file present in a training data folder.

Secondly in construction manual regex function used for preprocessing.

Q16. How did you do Data validation?

Ans: I have develop a manual regex function which will store the information related to the structure file of data as share by the mentor based on that only we will validate the data and perform operations.

Q17. How would you rate yourself in machine learning?

Ans: In the scale of 1-5 I rate myself at 3.5

Q18. What are the areas of machine learning algorithms that you already have explored?

Ans: I have explored LR, Logistic Reg, DT, Ensemble technique, KNN, Random forest, naïve bayes, and etc.

Q19.How did you optimize your solution?

Answer:

Well, model optimization depends on a lot of factors.

- Train with better data(increase the quality), or do data pre-processing steps more efficiently.
- Increase the quantity of data used for training.
- Increase the number of epochs for which the model was trained
- Tweak the batch input size, the number of hidden layers, the learning rate, rate of decay, etc. to produce the best results.
- Reduce Learning Rate to optimizes the model even further.

Q20. In which mode have you deployed your model?

Ans: In both local system only.

Q21. Have you ever used DVC, MLFLOW, and CICD?

Ans: Yes I have used in this project:

***CICD pipeline*** which is a method that builds, tests and deploys code by using automation and also help to automates your software delivery process.

***DevOps*** is also used in the model for efficiently compile, built and deploy their code to their production compute platform.

***MLFlow*** is also involved in this model development as to handle and manage the end to end machine learning life cycle.