Low Level Design (LLD)

Consumer Complaint Analysis (AIOPS PROJECT)

Revision Number: 1.0
Last date of revision: 19/02/2023

Shubham Kumar Chaturvedi

Document Version Control

Date Issued	Version	Description	Author
19th FEB 2023	1.0	First Draft	Shubham Kumar Chaturvedi

Contents

Document Version Control 2

Abstract 4

1 Introduction 5

1.1 Why this Low-Level Design Document? 5

1.2 Scope 6

1.3 Constraints 6

1.4 1.3 Risks 6
1.5 1.4 Out of Scope 7
2 Technical specifications 7
2.1 Predicting Disease 7
2.2 Logging 7
2.3 Database 7
3 Technology stack 8
4 Proposed Solution 9
5 Model training/validation workflow 10
6 User I/O workflow 12
7 Exceptional scenarios 13

Abstract

Complaint narratives are consumers' descriptions of their experiences in their own words. Consider what conclusions may be fairly drawn from reading consumers' descriptions of their experiences. We do not adopt their views or verify that their experiences are accurate or unbiased.

Complaints can give us insights into problems people are experiencing in the marketplace and help us regulate consumer financial products and services under existing federal consumer financial laws, enforce those laws judiciously, and educate and empower consumers to make informed financial decisions.

1 Introduction

1.1 Why this Low-Level Design Document?

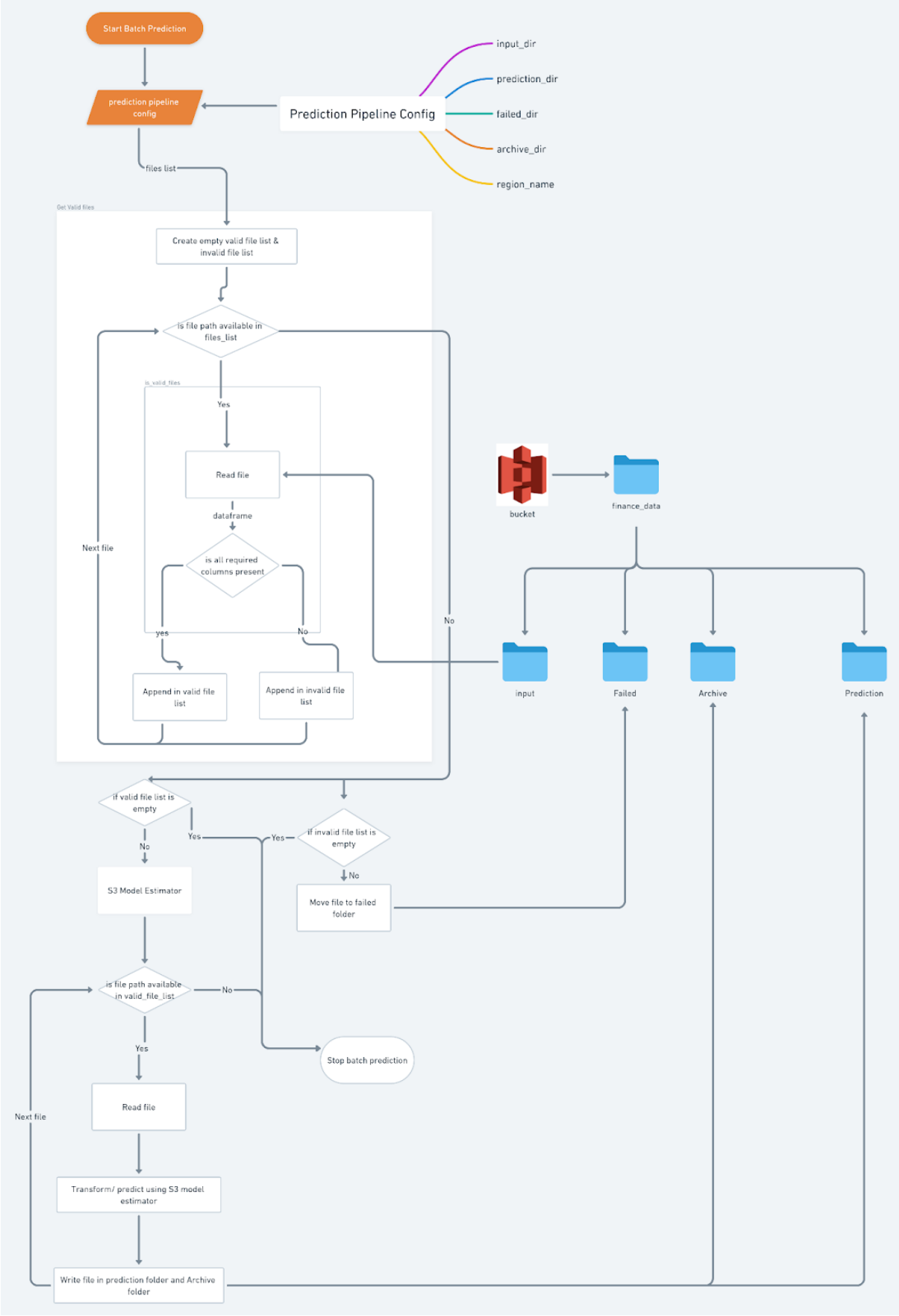
The purpose of this document is to present a detailed description of the Consumer complaints analysis system. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for

both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The main objective of the project is to predict if a person gives us insights into problems. People are experiencing the marketplace and help us regulate consumer financial products and services on dates,laws and educate and empower consumers to make informed financial decisions.

An Consumer complaint analysis (CCA) contains information, such as:

- Incorrect information on your report.
- Problem with a credit reporting company's investigation into an existing problem.
- Improper use of your reportImproper use of your report.
- Company response to consumer.
- Company public response.
- The date the CFPB sent the complaint to the company.
- How did the consumer submit the complaint to the CFPB?
- Only show complaints with narratives?


Start Batch Prediction
prediction pipeline config
Prediction Pipeline Config
input_dir
prediction_dir
failed_dir
archive_dir
region_name
files list
Get Valid files
Create empty valid file list & invalid file list
is file path available in files_list
Yes
is_valid_files
Read file
dataframe
is all required columns present
yes
No
Append in valid file list
Append in invalid file list
Next file
No
bucket
finance_data
input
Failed
Archive
Prediction
if valid file list is empty
Yes
No
S3 Model Estimator
if invalid file list is empty
Yes
No
Move file to failed folder
is file path available in valid_file_list
No
Yes
Stop batch prediction
Read file
Next file
Transform/ predict using S3 model estimator
Write file in prediction folder and Archive folder

This project shall be delivered in two phases:

Phase 1: All the functionalities with PyPi packages.

Phase2: Integration of UI to all the functionalities.

1.2 Scope

This software system will be an api solution and Web application. This system will be used to detect the useful complaints at earliest for better solution management, improved interventions, and more efficient care resource allocation using previous CCA records available. More specifically, Early detection of any preventable issue is important for better solution management. This system is designed to predict the legal and useful complaints from user complaints.

1.3 Constraints

We will only be selecting a few of the issues.

1.4 Risks

Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

2 Technical specifications

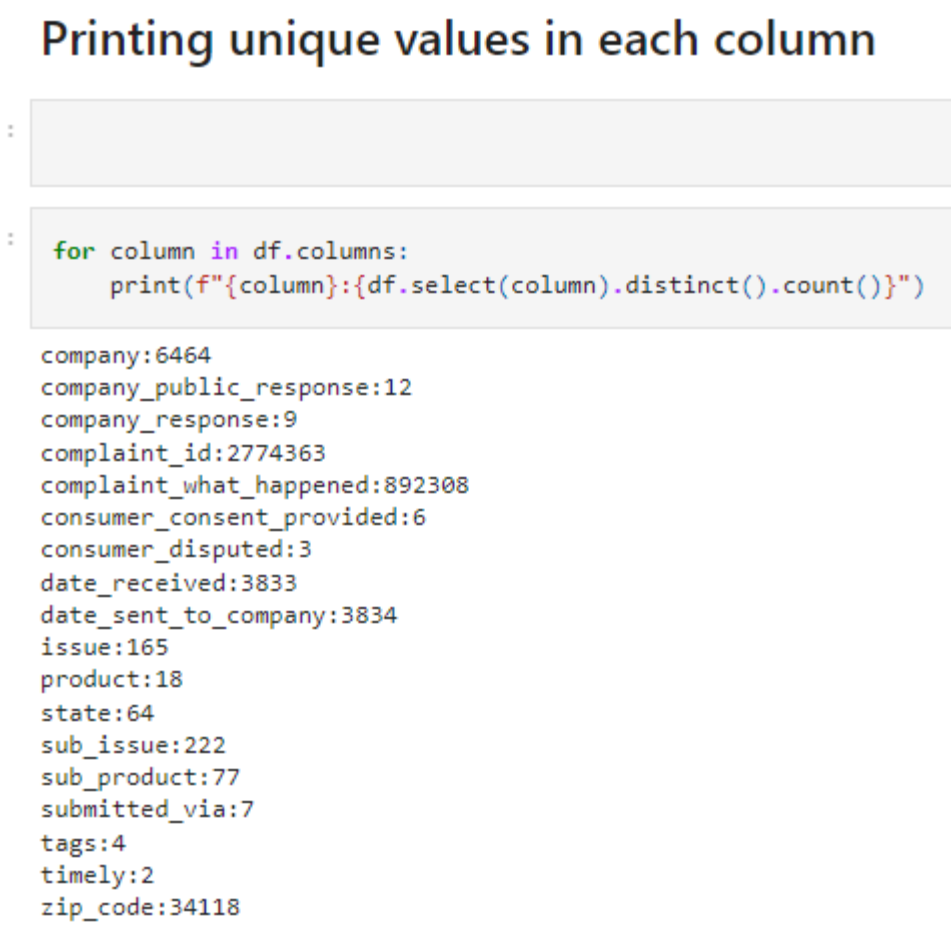
2.1 Dataset

Disease	Finalized	Source
Consumer Complaint Database	yes	https://www.consumerfinance.gov/data-research/consumer-complaints/search/?date_received_min=2011-12-01&page=1&searchField=all&size=25&sort=created_date_desc&tab=List

2.1.1 Consumer complaints dataset overview

There are a total of 2774363 complaint_id in the training set .

- Snapshot e



2.1.2 Input schema

Feature name	Datatype	Size	Null/Required
Consumer disputed	int	3	Required

2.2 Predicting

- The system presents the set of inputs required from the user.
- The user gives required information.
- The system should be able to predict whether information/issue for the chosen dispute immediately based on the user information.

2.3 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.
- System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.4 Database

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User input the issue.
2. The User gives required information.
3. The system stores each and every data given by the user or received on request to the database.

2.5 Deployment

1. AWS

3 Technology stack

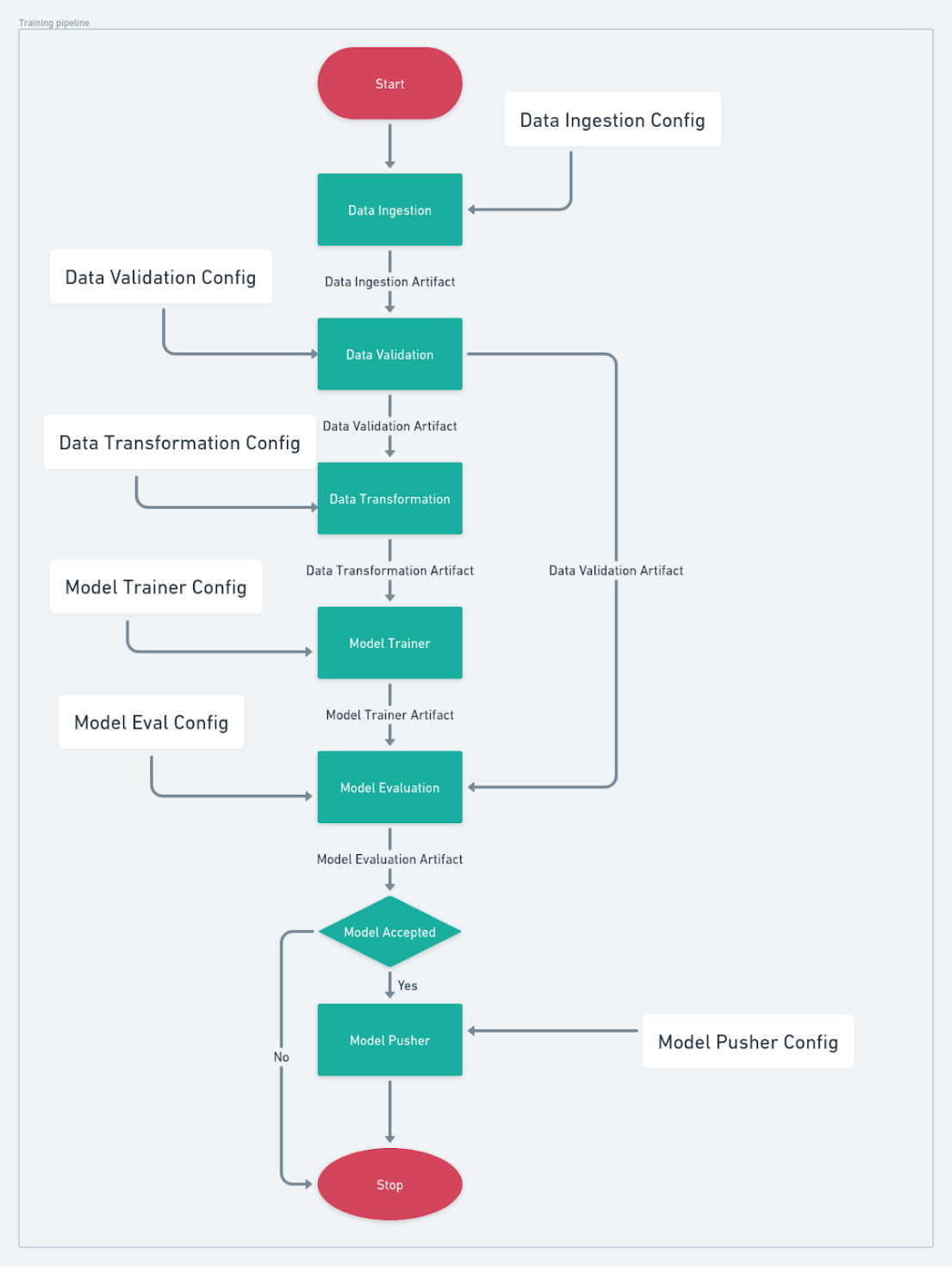
Front End	HTML/CSS/JS/React
Backend	Python,PySpark,PySpark ML Airflow as Scheduler.
Database	MongoDB
Deployment	AWS

4 Proposed Solution

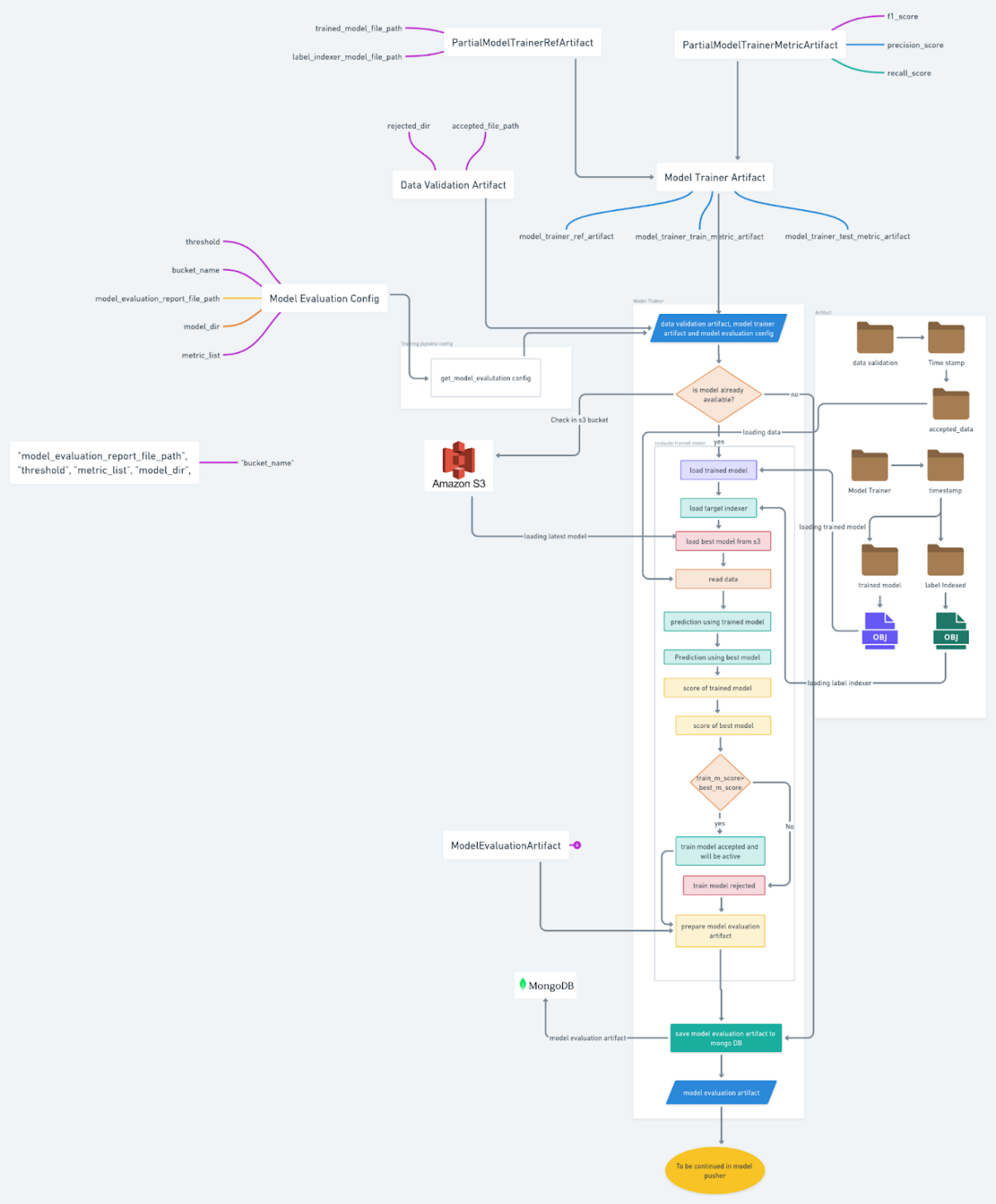
Based on the actual research paper, if we are using history of the complaints to predict the future then we might want to consider using LSTM. However, drawing a baseline in the form of some Machine Learning algorithm would be helpful. Why make a baseline model important? Well, to compare the performance of our actual model, let say LSTM in this case, is very important to ascertain that we are in the right direction as if performance of LSTM is not better than the baseline model then there is no point of using LSTM.

1. Baseline Model: Random forest classifier, since this is a classification problem.

5 Model training/validation workflow



6 User I/O workflow



7 Exceptional scenarios

Step	Exception	Mitigation	Module
19th FEB 2023	1.0	First Draft	Shubham Kumar Chaturvedi

8 Test cases

Test case	Steps to perform test case	Module	Pass/Fail

9 Key performance indicators (KPI)

- Time and workload reduction using the CCA model.
- Comparison of accuracy of model prediction and normal prediction.
- important steps have been taken.
- Use of automation in the finance sector.