PROJECT ARCHITECTURE

Project architecture

Why This Low Level Design

A low-level design document typically provides a detailed, in-depth description of the implementation of a system, focusing on specific aspects such as modules, classes, algorithms, data structures, and their interactions

Scope

The scope of the customer segmentation project encompasses several key areas that define the boundaries and extent of the work to be undertaken. It outlines the specific objectives, deliverables, and limitations of the project, providing a clear framework for what will be achieved.

The primary objective of this project is to develop a machine learning-based system for customer segmentation, enabling businesses to identify and categorize distinct customer groups based on their behaviors and characteristics. This involves collecting and preprocessing customer data, implementing various clustering algorithms, evaluating the performance of these algorithms, and interpreting the results to derive actionable insights. The ultimate goal is to provide businesses with a tool that helps in crafting targeted marketing strategies and improving customer satisfaction.

Constraints

This project on customer segmentation using machine learning faces several constraints that may impact its execution and outcomes. Identifying and understanding these constraints is essential for setting realistic expectations and mitigating potential challenges. The quality and availability of data are significant constraints for this project. The effectiveness of customer segmentation relies heavily on the accuracy, completeness, and relevance of the data collected. Issues such as missing values, inconsistent data formats, and insufficient data can hinder the preprocessing and analysis stages. The project requires substantial computational resources for data processing, model training, and evaluation. The complexity and volume of customer data, combined with the need to run multiple clustering algorithms, can demand significant processing power and memory.

Risks

The customer segmentation project using machine learning involves several risks that can potentially impact its success and effectiveness. Identifying these risks early on allows for the development of mitigation strategies to address them. Handling sensitive customer data poses significant privacy and security risks. Unauthorized access, data breaches, or misuse of customer information can lead to legal repercussions and damage the organization's reputation. Ensuring robust data protection measures, such as encryption and secure access controls, is crucial to mitigate these risks.

Out of Scope

To clearly define the boundaries of the customer segmentation project, it is essential to identify aspects that are beyond its scope. This helps manage expectations and focus resources on the primary objectives. The project will not address the implementation of real-time data processing or streaming analytics. While customer segmentation could benefit from real-time data updates, this project is limited to batch processing of historical data. Real-time processing requires additional infrastructure and technologies that are not covered within the current scope.

## Customer Segmentation Datasets Overview

```
1    ID,Gender,Ever_Married,Age,Graduated,Profession,Work_Experience,Spending_Score,Family_Size,Var_1,Segmentation
10247    467451,Female,Yes,51,Yes,Artist,0.0,Low,1.0,Cat_1,C
10248    467559,Male,Yes,58,Yes,Executive,0.0,Average,4.0,Cat_6,A
10249    467727,Male,Yes,53,Yes,Artist,7.0,High,4.0,Cat_3,C
10250    467789,Male,No,20,No,Healthcare,1.0,Low,5.0,Cat_6,B
10251    467828,Female,Yes,25,Yes,Artist,0.0,Average,2.0,Cat_2,B
10252    467882,Female,No,39,Yes,Artist,1.0,Low,1.0,Cat_6,D
10253    467888,Male,No,29,No,Doctor,0.0,Low,4.0,Cat_6,D
10254    467904,Female,No,31,Yes,Artist,14.0,Low,1.0,Cat_6,D
10255    465022,Female,Yes,71,Yes,Lawyer,0.0,High,2.0,Cat_6,C
10256    465087,Male,No,29,Yes,Doctor,0.0,Low,4.0,Cat_6,A
10257    465105,Male,Yes,65,No,Executive,8.0,Average,2.0,Cat_6,C
10258    465122,Female,Yes,41,Yes,Artist,3.0,Average,2.0,Cat_6,A
10259    465172,Female,Yes,43,Yes,Artist,8.0,Average,2.0,Cat_6,D
10260    465181,Female,No,25,Yes,Healthcare,8.0,Low,1.0,Cat_4,D
10261    465182,Female,No,73,Yes,Lawyer,1.0,Low,1.0,Cat_6,A
10262    465196,Female,No,31,Yes,Artist,2.0,Low,2.0,Cat_4,D
10263    465203,Male,No,29,Yes,Healthcare,3.0,Low,3.0,Cat_6,D
10264    465372,Female,No,40,Yes,Entertainment,0.0,Low,1.0,Cat_6,A
10265    465410,Male,No,30,No,Healthcare,0.0,Low,4.0,Cat_5,D
10266    465461,Male,Yes,69,No,Entertainment,1.0,Average,3.0,Cat_6,D
10267    465598,Male,No,27,Yes,Healthcare,6.0,Low,5.0,Cat_2,C
10268    465655,Female,No,49,Yes,Artist,0.0,Low,1.0,Cat_6,C
10269    465727,Male,No,40,No,Marketing,0.0,Low,5.0,Cat_4,B
10270    465733,Male,Yes,32,Yes,Entertainment,1.0,Average,2.0,Cat_4,B
10271    465751,Female,Yes,46,Yes,Artist,9.0,Low,2.0,Cat_6,A
10272    465866,Male,Yes,42,Yes,Artist,6.0,Low,3.0,Cat_6,A
10273    465953,Female,No,51,Yes,Artist,1.0,Low,1.0,Cat_6,C
10274    465974,Male,Yes,45,Yes,Executive,5.0,High,4.0,Cat_4,A
10275    465997,Female,No,31,No,Doctor,0.0,Low,3.0,Cat_6,C
10276    466007,Male,Yes,61,No,Artist,0.0,Low,1.0,Cat_7,A
10277    466154,Female,Yes,51,Yes,Artist,1.0,Average,5.0,Cat_3,C
10278    466156,Female,Yes,48,Yes,Doctor,8.0,Average,3.0,Cat_2,B
10279    466157,Male,Yes,72,Yes,Homemaker,9.0,Low,,,D
10280    466208,Male,Yes,69,No,Executive,2.0,High,1.0,Cat_6,D
10281    466210,Female,Yes,58,Yes,Artist,1.0,High,4.0,Cat_6,C
```

Dependent and Independent Schema

Features Name	DataTypes
Gender	object
Ever married	object
Age	int
Graduated	object
Profession	object
Work Experiences	object
Spending score	object
Family size	int

Data Ingestion WorkFlow
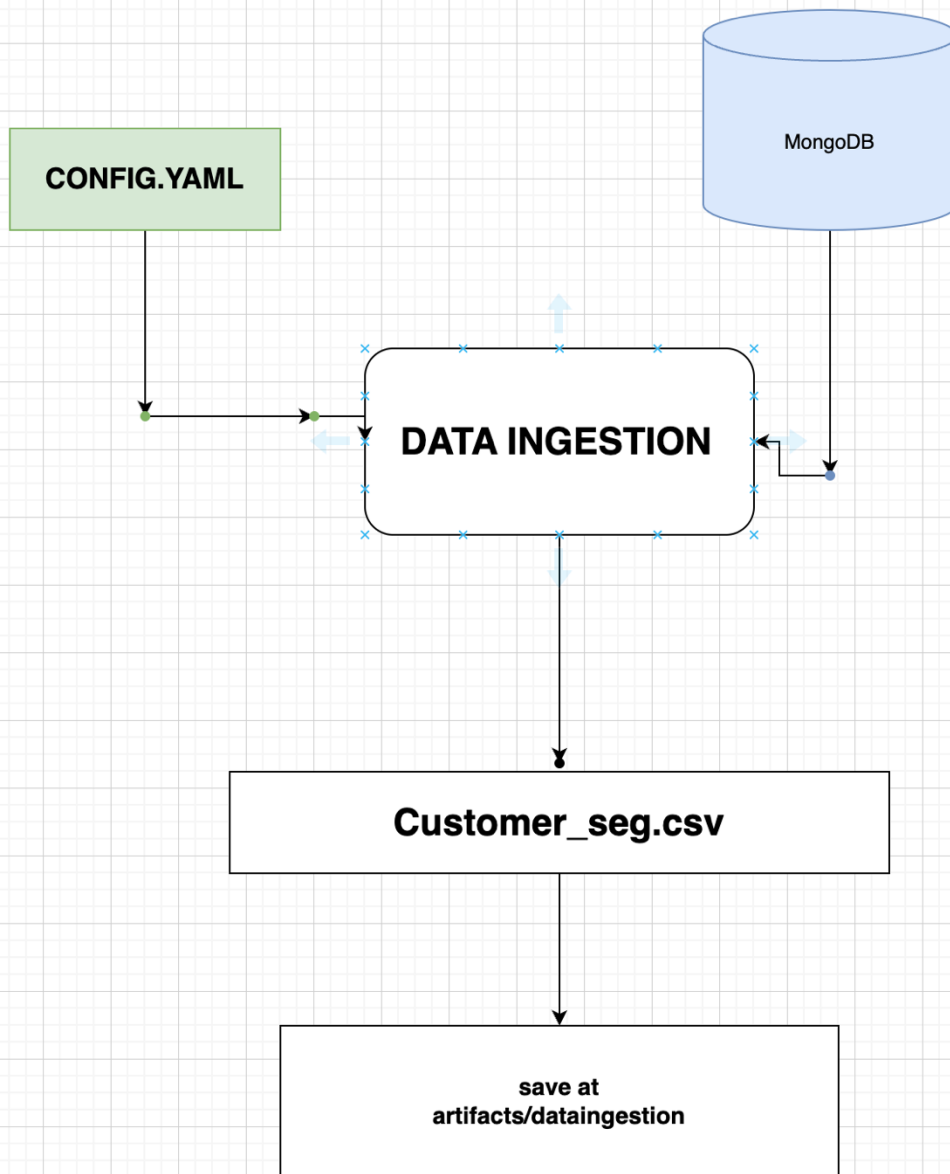
Data Validation WorkFlow
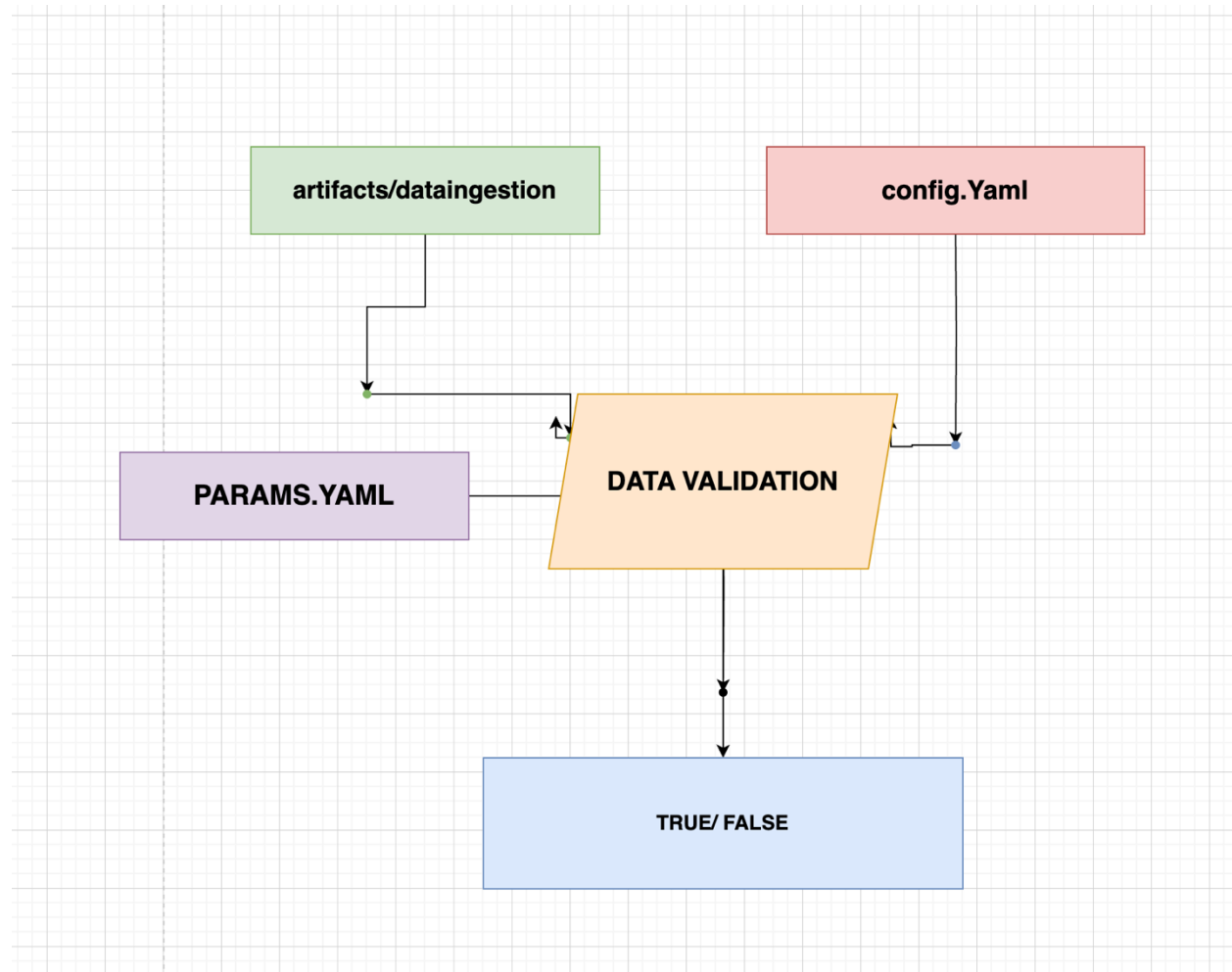
Data Transformation WorkFlow

```
          ┌────────────────────────┐              ┌────────────────────────┐
          │  artifacts/dataingestion│              │      config.Yaml        │
          └────────────────────────┘              └────────────────────────┘
                      │                                         │
                      ▼                                         ▼
          ┌─────────────────────────────────────────┐
          │                                          │
  ┌──────────────┐      ╱ Data Transformation ╲
  │ PARAMS.YAML  │─────▶╲                      ╱
  └──────────────┘        ╲_____╱
                                   │
                                   ▼
                        ┌────────────────────┐
                        │   Data Cleaning     │
                        └────────────────────┘
                                   │
                                   ▼
                    ┌──────────────────────────┐        ┌──────────────────────────────┐
                    │ Feature Engineering      │───────▶│ Save to                      │
                    │ Pipeline                 │        │ artifact/datatransfromation  │
                    └──────────────────────────┘        │ Preprocessor.pkl             │
                                   │                     └──────────────────────────────┘
                                   ▼
                        ┌────────────────────┐
                        │ Perprocessed_Data.csv│
                        └────────────────────┘
```
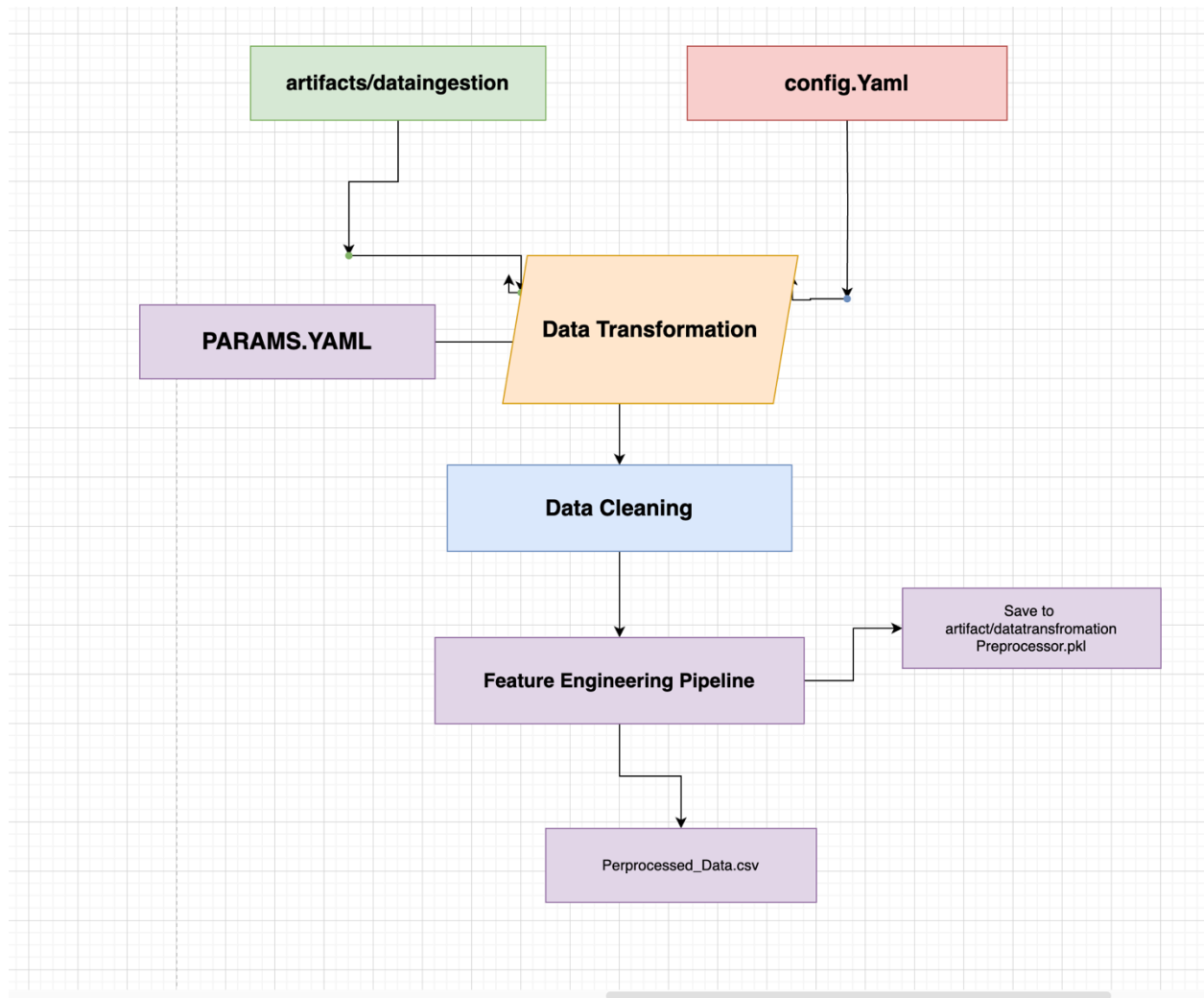
ModelTraining WorkFlow

```
┌─────────────────────────┐          ┌─────────────────────────┐
│  artifacts/dataingestion │          │       config.Yaml        │
└─────────────────────────┘          └─────────────────────────┘
                    │                              │
                    │                              │
                    ▼                              ▼
┌──────────────────┐   ┌───────────────────────┐
│  PARAMS.YAML     │──▶│    MODEL TRAINING      │
└──────────────────┘   └───────────────────────┘
                                   │
                                   ▼
                       ┌───────────────────────┐
                       │          PCA           │
                       └───────────────────────┘
                                   │
                                   ▼
                       ┌───────────────────────┐
                       │    Kmean clustering    │
                       └───────────────────────┘
                                   │
                                   ▼
                       ┌───────────────────────┐
                       │        Save to         │
                       │  artifact/model_trainer │
                       │        model.pkl        │
                       └───────────────────────┘
```