Flight Price Prediction System

Spring 2022 CSYE 7200

Luo Chen

Project Description

- Project: Flight Price Prediction System

  A system used to predict the price trend of a flight

- Team Members:

Yuhan Yang	1094267
Luo Chen	1564677
Kang Shentu	1569432

Project Description

- Project: Flight Price Prediction System

  A system used to predict the price trend of a flight

- Factors
  - Different airline
  - Days before departure
  - Departure/Arrival time
  - Source/Destination city
  - Economy/Business class

Project Description

- Use Cases
  - To Customers
    - find the best flight
    - when to purchase the air ticket
  - To Business
    - recommend flights for customers

Methodology

- Algorithm
- Engineering

Methodology

- Algorithm
  - Linear Regression
  - XGBoost (Cost-effective)
  - Transformer

Methodology

- Engineering
    - Spark
        - Data Processing
        - Training
        - Inferring
    - Java
        - Web Service
    - Others
        - Operation
        - Docker

Methodology

- Engineering
  - Usability
  - Accuracy
  - Reliability

Methodology

- Engineering
  - Usability

    A system can provide services for customers to use
    - Offline Learning
      - Load preprocess data
      - Train
      - Save Model
    - Static Predicting (synchronous/asynchronous/streaming)
      - Load Model
      - Receive and Transform data
      - Predict
      - Output

Methodology

- Engineering
  - Accuracy

    A system can revise the model by feeding new data
    - Data crawling
    - Online Learning
      - Feed data
      - Process data
      - Update parameters or train a new Model
      - Save Model
    - Static Predicting
      - Load updated model time by time

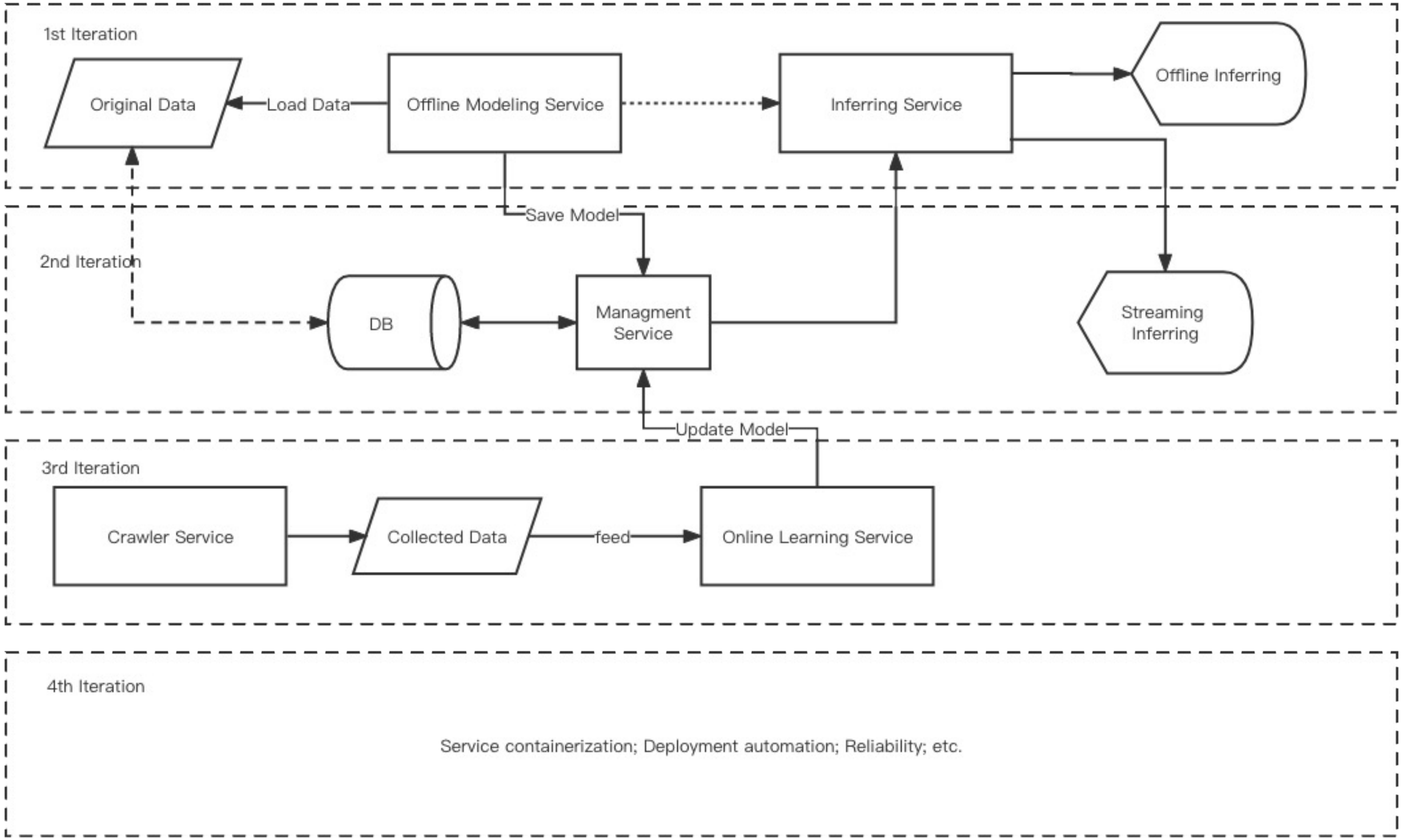
Methodology

- Engineering
  - Reliability

    A System needs high availability and ease for operations

    - Service containerization
    - Auto Deployment
    - Different components
    - Etc.

Architecture



Architecture

- Repositories
  - Web Crawler (In Scala)
  - Online Learning (In Scala)
  - Streaming Predicting (In Scala)
  - Management Service (In Java)

These repositories will be pushed onto GitHub.

Data Source

Original Dataset:

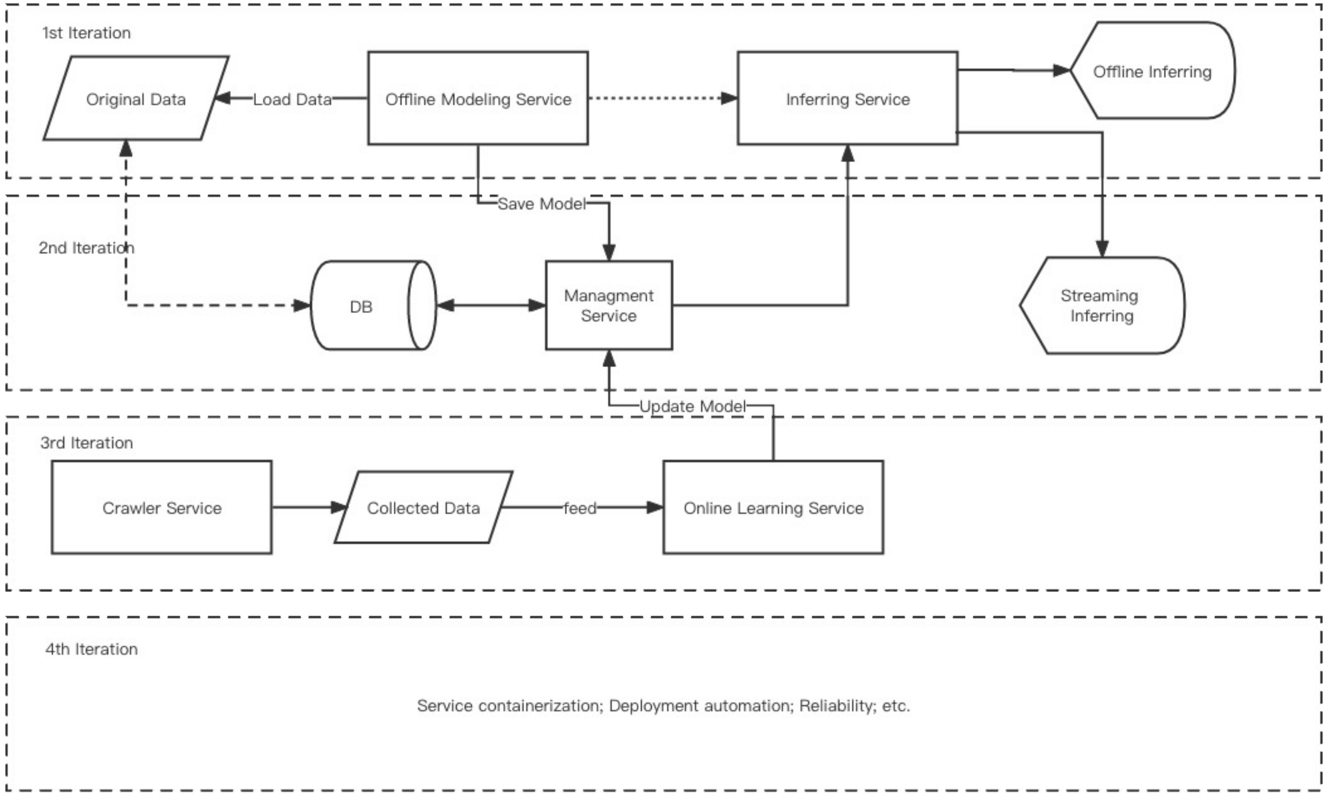
Kaggle: https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction

Dataset contains information about flight booking options from the website EaseMyTrip for flight travel between India's top 6 metro cities.
There are 300261 datapoints and 11 features in the cleaned dataset.

Following Data:

Crawl from websites like EaseMyTrip, SkyScanner, etc.

Milestones



1st Week: Implement the basic system to perform offline training and batch predicting.

2nd Week: Add the service to manage data, models and predictions. Implement streaming processing for inferring service.

3rd Week: Complete crawler service. Update offline learning To online learning. Implement the workflow for the while system.

4th Week: Optional work. Strengthen reliability of our system.

Acceptance Criteria

- The response time of the API for prediction for one input is less than 1s
- Training time of static model (offline training) should less than 1 hour
- Updating model by new data retrieving from web-crawler every 2 hours
- The R2 score for the regression model should be higher than 0.70

Goals of the project

- Help us understand the characteristic of Scala and advantages of Spark.
- Learn to design and implement a big data system.
- Develop the ability to work with our teammates.
- Learn how to use machine learning to solve problems in real life.

Thank You!

Spring 2022 CSYE 7200

Luo Chen