

AST Solution Technical Assessment

2) Write a function that removes duplicate characters from string. Provide at least 3 solutions. Which is best in your opinion? Why?

Solution 1:

```
# @profile
def solution1(value: str) -> str:
    length_value = len(value)
    dummy_value = ""

    for pos in range(length_value):
        if value[pos] not in dummy_value:
            dummy_value = dummy_value.__add__(value[pos])
    return dummy_value
```

Solution 2:

```
# @profile
def solution2(value: str) -> str:
    return "".join(sorted(set(value), key=value.index))
```

Solution 3:

```
# @profile
def solution3(value: str) -> str:
    return "".join(OrderedDict.fromkeys(value))
```

Solution	Ordered	Time Complexity
1	✓	$O(n)$
2	✓	$O(n \log n)$
3	✓	$O(\log n)$

Solution 3 is fastest approach as it takes less time compute and preserving order of characters that is present original input string.

5) Propose a birds-eye view technical design for a ML based program that reads vehicles license plates and track their movement based on input from multiple cameras. Consider in your solution:

- a. Data handling
- b. Scaling and optimization
- c. Security
- d. Libraries

Solution: In this solution I'm only assuming that the requirement is to read vehicle license plate and track the vehicle movement using vehicle license plate from multiple cameras.

Below are the block diagrams (See Figure1, Figure2, Figure3) for designing ML application

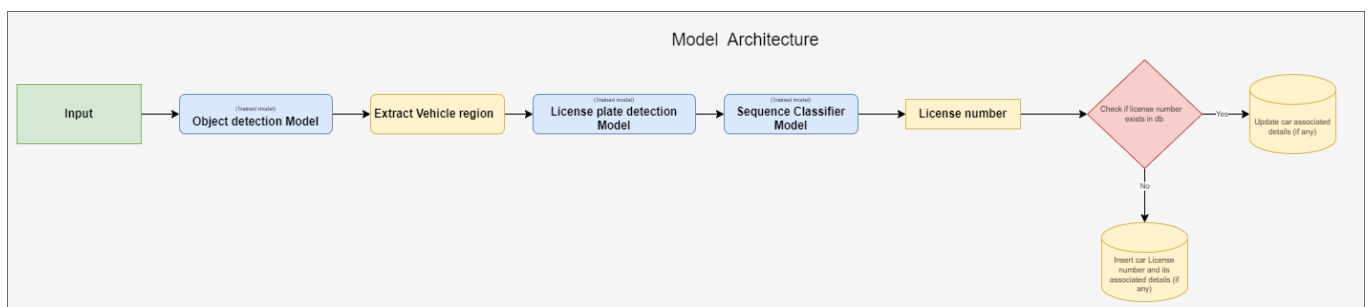


Figure 1: Model Architecture

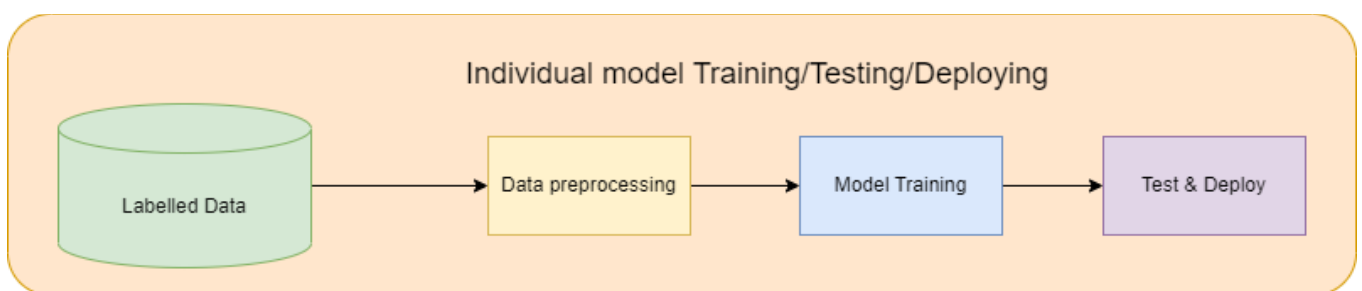


Figure 2: Individual Model Training/Testing/Deploying

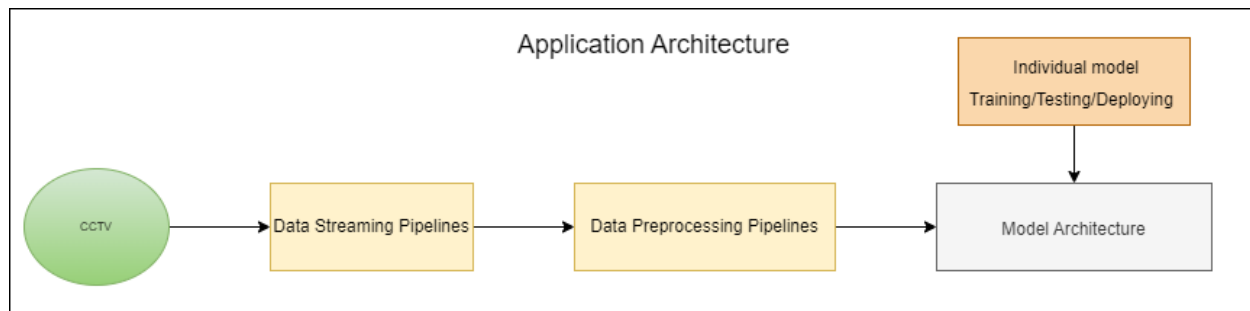


Figure 3: Application Architecture

Libraries	Usage
PyTorch/Tensorflow/Keras	Model building, parameter tuning
ML flow	For tracking and logging model details
Apache Spark ML	ETL pipelines
Apache Airflow	For tuning and scaling ML models
Amazon cloud infrastructure (S3, EC2, SageMaker) (or any cloud platform provider)	For storage, performing computation, security privilege, data streaming
ML in python libraries	Hyperopt, OpenCV

Table 1: Libraries

There are totally three models (Object detection, License plate detection model and Image to text (sequence classifier model) needed to get license number from a vehicle captured on CCTV camera (see Figure1. All these models can be built from scratch or fine tuned on existed pre trained models.

For object detection we can use, any start of art technique like Mask RCNN, YOLO to identify vehicle regions. And a trained CNN model (License plate detection model) is used to extract the license plate from the identified regions. And its output is given to a trained image to text classifier to get license number of that vehicle in the identified region. Once vehicle license number is extracted, it is stored and update in database (so we can track the vehicles).

In training and deploying, we employ data preprocessing airflow pipelines with Apache spark on SageMaker for fine tuning and scaling the models using single machine training technique and distributed training. Once the entire Model architecture is built and deployed, we stream CCTV data through S3 buckets and feed it to Model architecture through data processing airflow pipelines. And security of entire model is secured as we are dependent on cloud infrastructure with proper security privilege constraints.