

# HUMAN ACTIVITY DETECTION IN LOGISTICS

R GOKULAVASAN  
(1833014)  
**Guided by,**  
Dr. Marimuthu

HAPPYMONK.AI

## HAPPYMONK.AI



This company specializes in connecting people and devices through the use of AI and privacy-preserving techniques. They develop products and solutions in the field of emerging technologies such as autonomous machines, cyber-physical systems, and command & control centres. Their team of data scientists and AI engineers is led by an experienced Applied AI Master, Dr. Snehanshu Saha.

They focus on creating state-of-the-art methods powered by deep learning, computer vision, and cryptography to assist businesses and individuals in adopting blockchain-driven technologies and transforming traditional lives.

# Problem Statement

01

The problem in the company is workers mishandling products in the warehouse, which can result in damaged goods, increased costs, and customer dissatisfaction. Workers tend to sit on, stand on, and throw packages without any care, leading to a high rate of product mishandling.

02

This issue is affecting the efficiency of the warehouse operations and can have a negative impact on the reputation of the logistics company. A solution is needed to identify workers' activities in real-time and reduce product mishandling.



# Objective

- 01 The mishandling of products by workers in the warehouse of a logistics company can lead to damaged goods, increased costs, and customer dissatisfaction. To address this issue, a machine learning-based activity recognition system has been proposed.
- 02 The system aims to recognize workers' activities in real-time, such as carrying, walking, sitting, standing, and throwing, using multi-camera support and achieve high accuracy. The proposed system is expected to reduce product mishandling and improve warehouse efficiency.



# Tools



SKlearn  
For Analysis



Pytorch - Slowfast  
For Model Training



Pandas and Numpy  
Basic Data Analysis



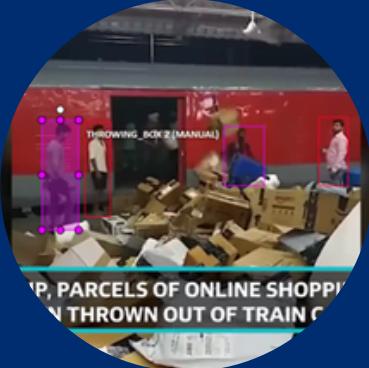
Cvat & Jupyter Notebook  
Data Preprocessing

# Data

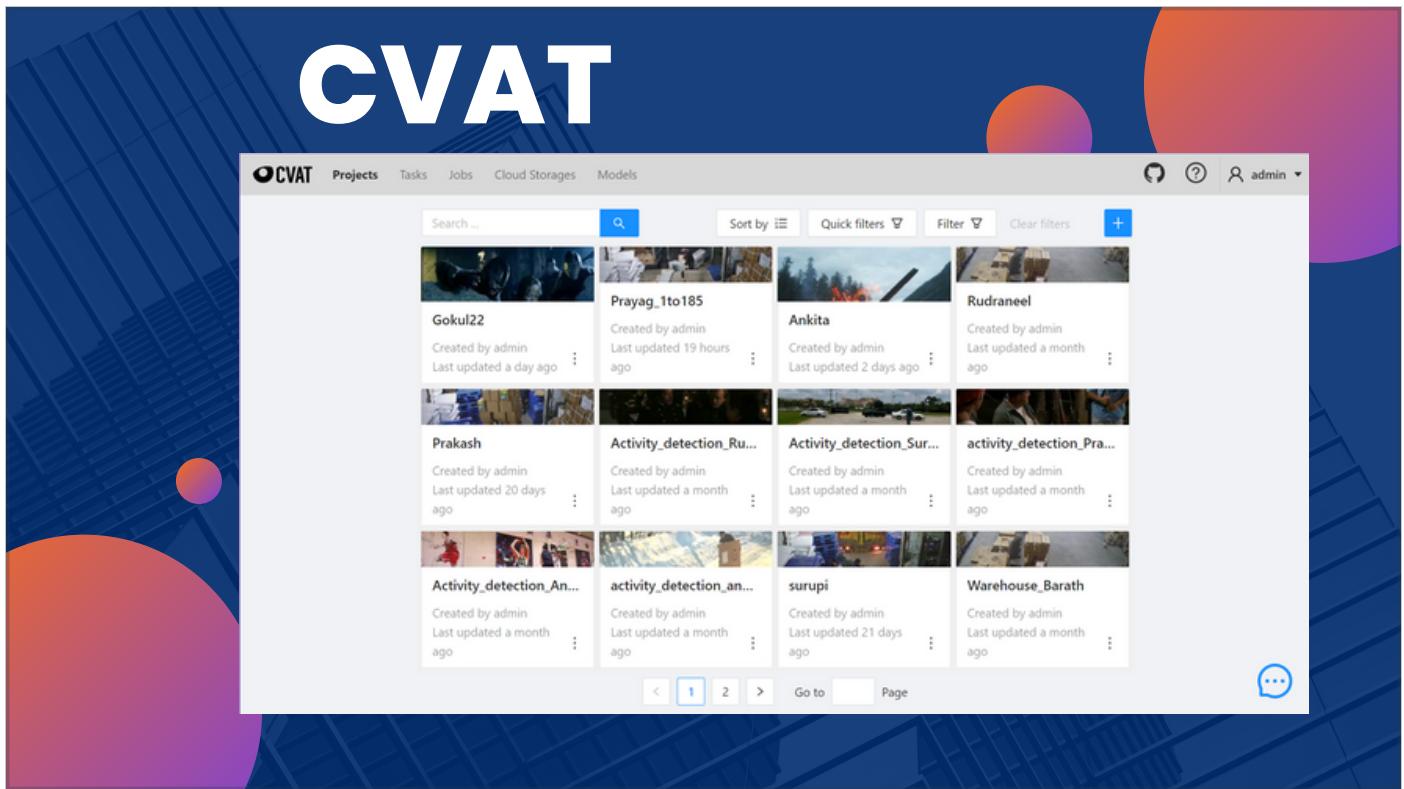
Vid_Name	Frame_Nc	Timestamp	Annotatio	Annotatio	Annotatio	Annotatio	Class
sample_vid185.mp4	0	0	0.381907	0.670576	0.177114	0.34985	2
sample_vid185.mp4	29	0.966667	0.380813	0.67641	0.179302	0.361517	2
sample_vid185.mp4	59	1.966667	0.38916	0.672526	0.167969	0.371094	2
sample_vid185.mp4	89	2.966667	0.389404	0.668294	0.17334	0.35612	2
sample_vid185.mp4	119	3.966667	0.389404	0.671549	0.168457	0.360026	2
sample_vid185.mp4	149	4.966667	0.395996	0.669596	0.146484	0.366536	2
sample_vid185.mp4	179	5.966667	0.407471	0.67513	0.162598	0.372396	2
sample_vid185.mp4	209	6.966667	0.373022	0.66569	0.189404	0.352214	2
sample_vid185.mp4	239	7.966667	0.37793	0.672526	0.162109	0.401042	2
sample_vid185.mp4	269	8.966667	0.381787	0.666016	0.17373	0.385417	2
sample_vid185.mp4	299	9.966667	0.396484	0.67513	0.148438	0.364583	2
sample_vid184.mp4	0	0	0.532959	0.701172	0.272949	0.39974	2
sample_vid184.mp4	29	0.966667	0.507568	0.685221	0.17627	0.404297	2
sample_vid184.mp4	59	1.966667	0.521484	0.701497	0.209961	0.427734	2
sample_vid184.mp4	89	2.966667	0.507812	0.704102	0.168945	0.426432	2
sample_vid184.mp4	119	3.966667	0.525671	0.651624	0.154155	0.521868	3
sample_vid184.mp4	149	4.966667	0.561206	0.661829	0.192422	0.530612	3
sample_vid184.mp4	179	5.966667	0.551911	0.680781	0.173833	0.527695	3
sample_vid184.mp4	209	6.966667	0.451877	0.62903	0.131196	0.409622	3
sample_vid184.mp4	239	7.966667	0.428369	0.656729	0.138848	0.368809	3
sample_vid184.mp4	269	8.966667	0.390137	0.666341	0.150391	0.367839	2
sample_vid184.mp4	299	9.966667	0.388916	0.671224	0.17334	0.375	2
sample_vid184.mp4	0	0	0.50709	0.700061	0.158601	0.422526	2

- The data we needed for this project is footages of those 5 activities
- The data is collected from UCF dataset, AVA
  - Kinetics dataset,
  - Behave dataset, pexels
  - and from Youtube.

## Video Data



Review 2 Page 7 of 16



# Preprocessing

01

Annotating only required  
frames in a video clip

02

Removing null frames

03

Converting annotated txt  
files to csv format

## Data Collection

Collected 1000 video clips for the 5 activities over different open sources.

## Pre-processing

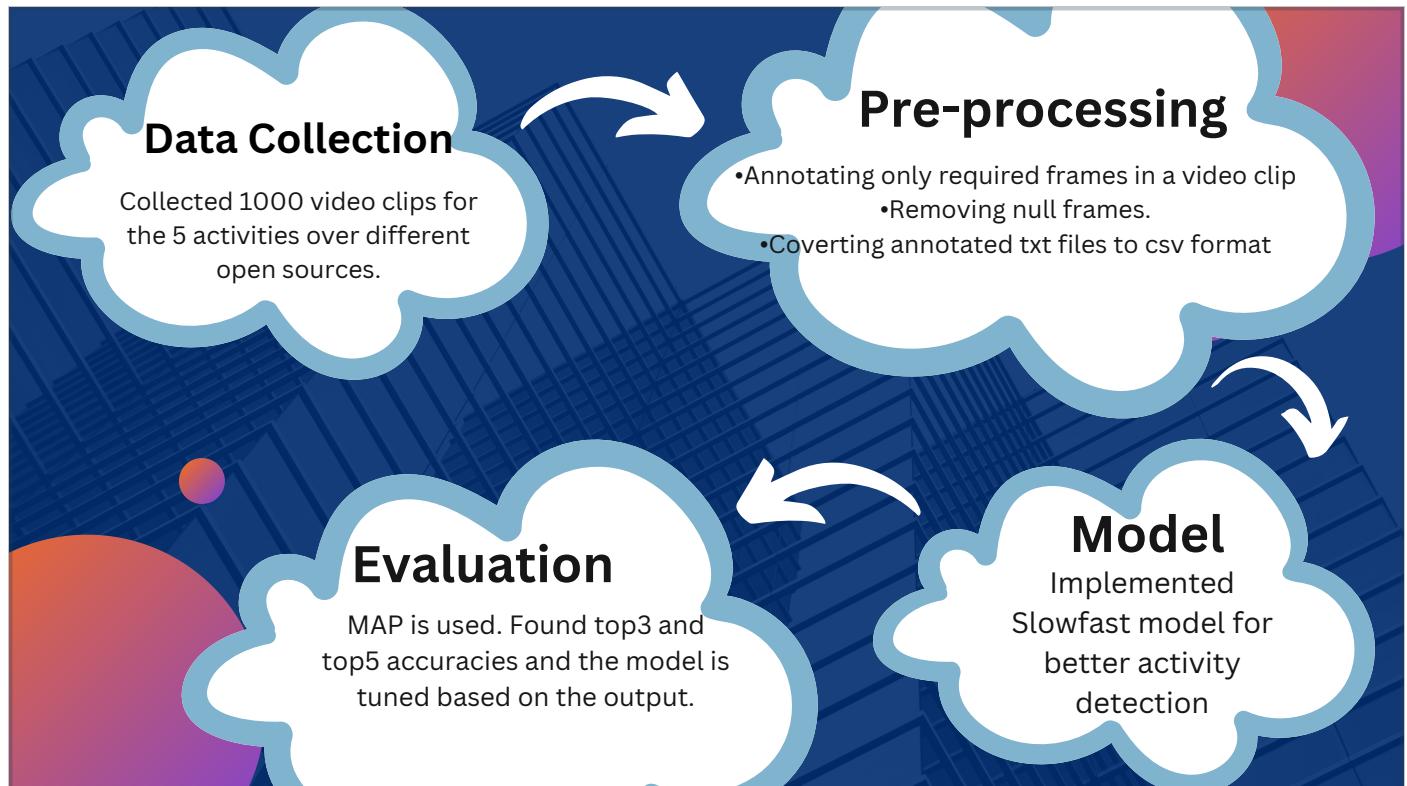
- Annotating only required frames in a video clip
- Removing null frames.
- Converting annotated txt files to csv format

## Evaluation

MAP is used. Found top3 and top5 accuracies and the model is tuned based on the output.

## Model

Implemented Slowfast model for better activity detection

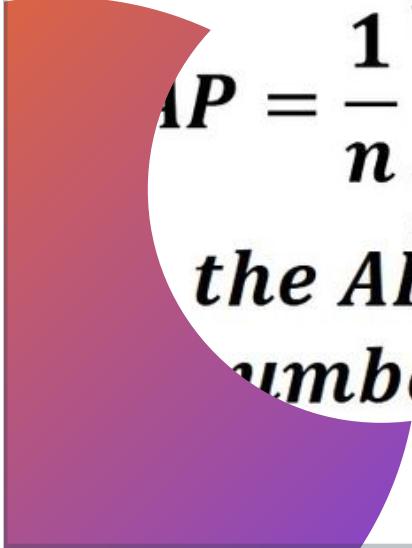




## Models

- LSTM
- Resnet
- Slowfast

- Why Slowfast?
- Spatial
- Temporal


$$AP = \frac{1}{n} \sum_{k=1}^{k=n} A_k$$

*the AP of each number*

## Evaluation metrics

•MAP (Mean Average Precision) is a commonly used evaluation metric in machine learning and information retrieval. It measures the average precision of a model over multiple classes or categories.

- 1.Top-3 accuracy: This measures the percentage of instances where the correct class is among the model's top 3 predictions.
- 2.Top-5 accuracy: This measures the percentage of instances where the correct class is among the model's top 5 predictions.

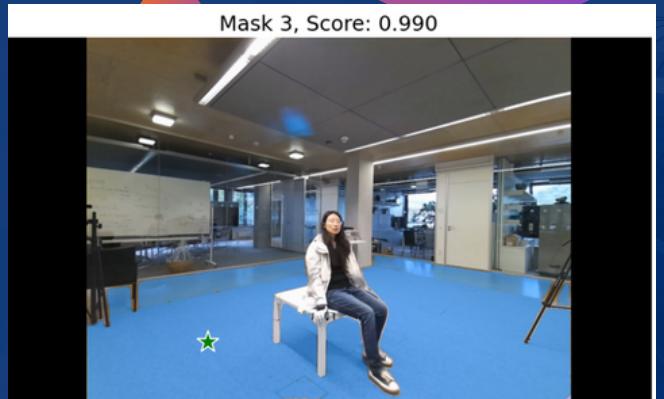
## Conclusion

•MAP - 40 %

•Recall- 91% Precision - 32%

*The model has attained much low performance for now. After adding more data and tuning of Parameters the model will perform well.*

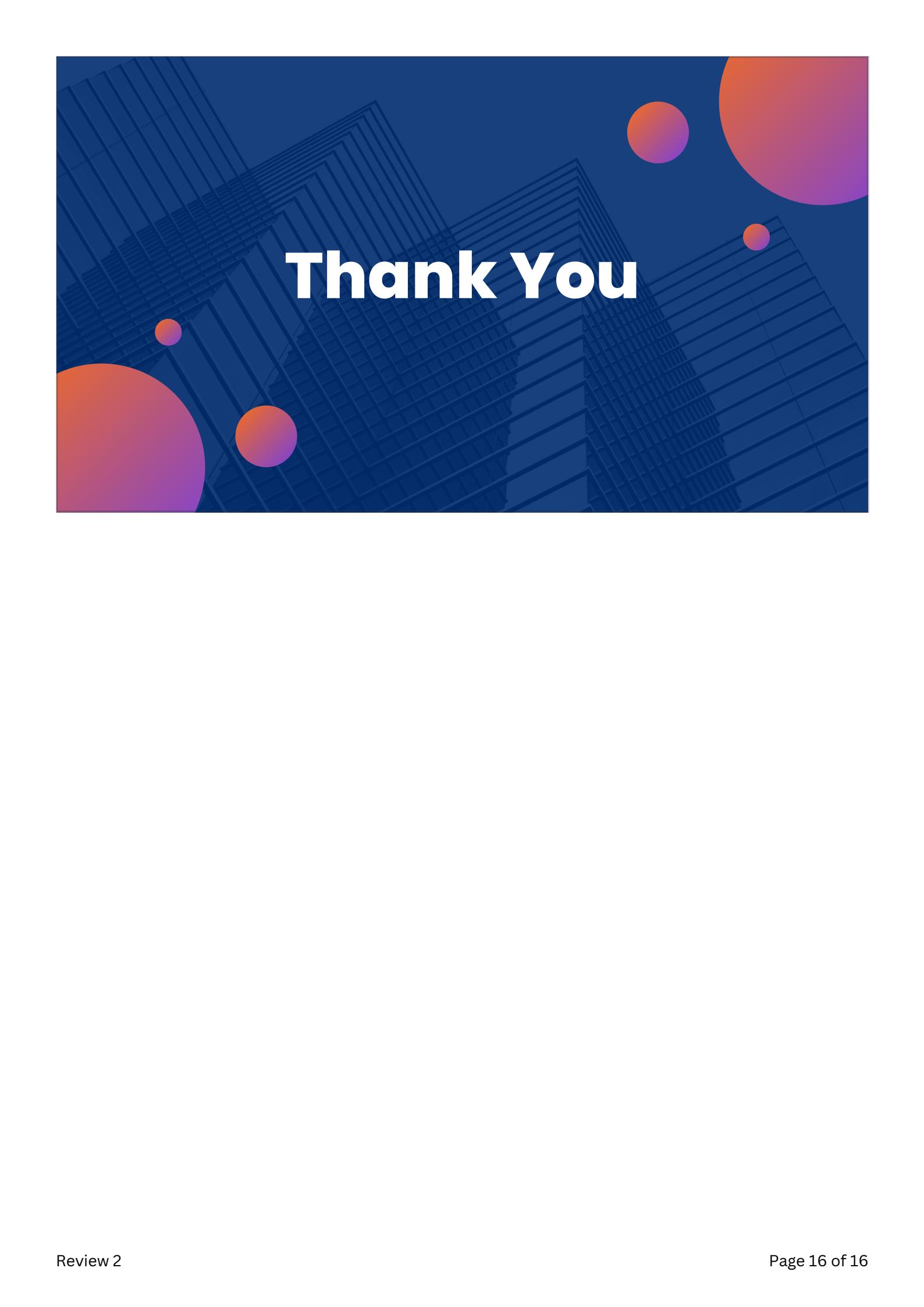
# SAM



## Future Work

- Integration of SAM
- After data collection completion more work on hyperparameters will be done
- More work will be done to increase the MAP of the model.





# Thank You