ET610
Learning Analytics

Course Project

Bhurle Mayur Ramesh	23M1027
Amalendhu SP	23D1204
Param Shah	200110079

Google Colab Link

https://colab.research.google.com/drive/1Y9AMFARZ2XvwTOLHvmRpCQnn30hAR99Z?usp=sharing

Author: Mayur

Links to all dataset:

Link of output given by ML Model for the videos is on slide number 8 and 9.

Link for Extracted frames of "Sample Video"= https://drive.google.com/drive/folders/130PWYxyFyVAOcWpGpd4PzgPodLRx88zR?usp=sharing

Link for Extracted Frames of video "v1" = https://drive.google.com/drive/folders/1Vt3vks_gPGU42r5Yw9xQGIZiTNmilWOp?usp=sharing

Link for Extracted Frames of video "v2" = https://drive.google.com/drive/folders/1mRE-V4bHFvL7R6CAL0KduLRfxbZLdKOG?usp=sharing

Link for the video data "v1" and "v2"= https://drive.google.com/drive/folders/1ZQzUvfL6VNSZrM0mAfEDheyl-4T7EYa8

Link for sample video= https://drive.google.com/file/d/1jXxgI-HKP8OH02PZf-wsJUsaX1QeWxjl/view

Author: Mayur

PROBLEM STATEMENT

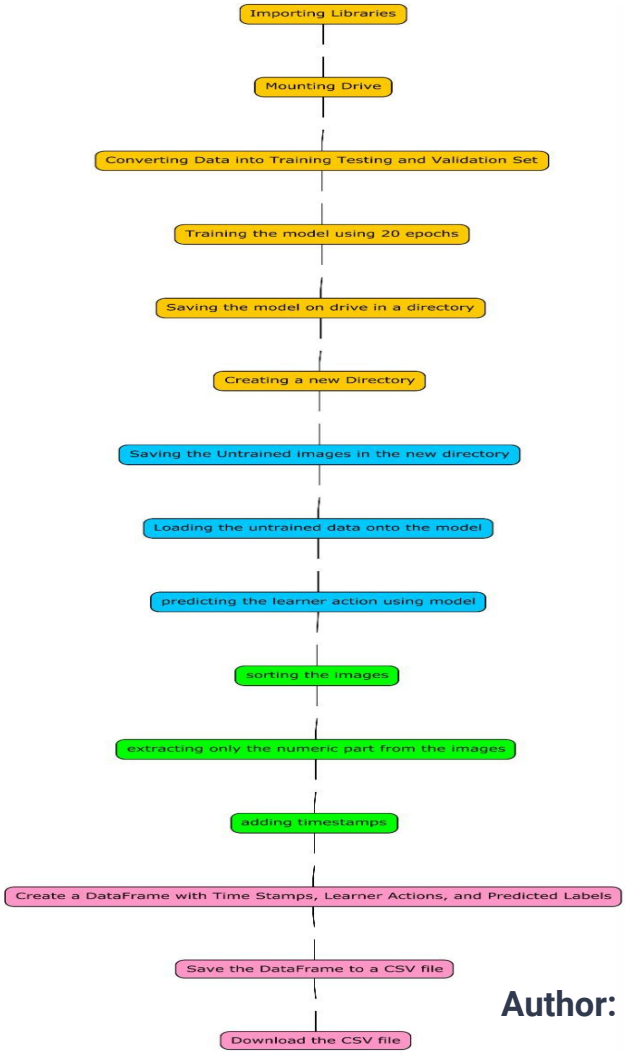
Develop an algorithm to automatically classify the web pages into majorly four different environments(ChatGPT, Jupyter, Excel and Other) and identify the corresponding learner action.

Data Description

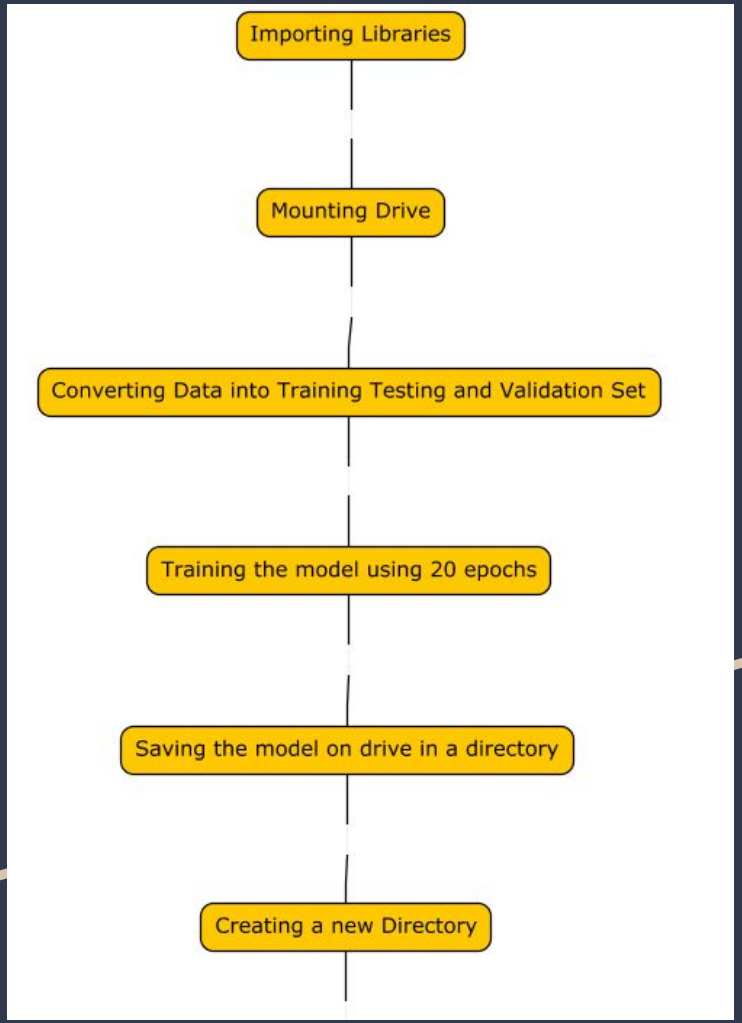
A Screen Recording of a Novice Programmer attempting to solve a given problem statement using ChatGPT, Jupyter, Google Search, Excel etc

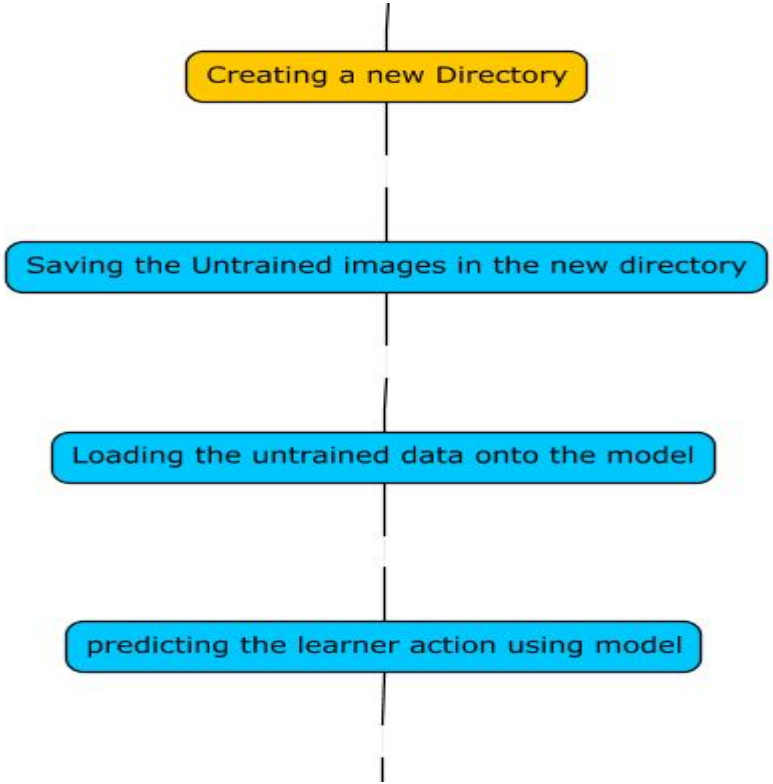
Author: Mayur

Model Flowchart

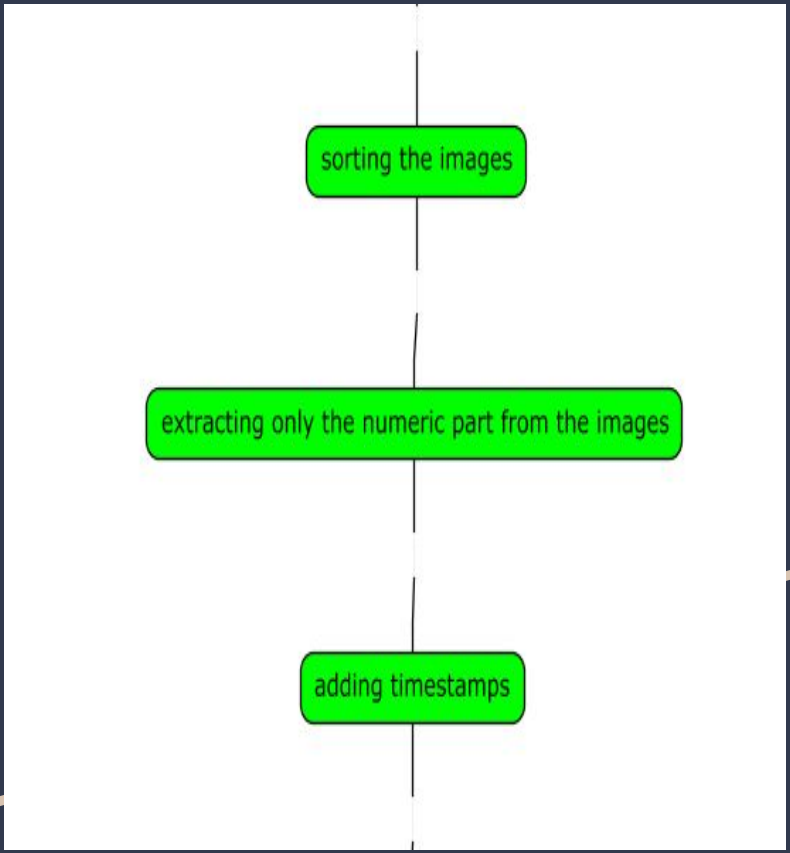


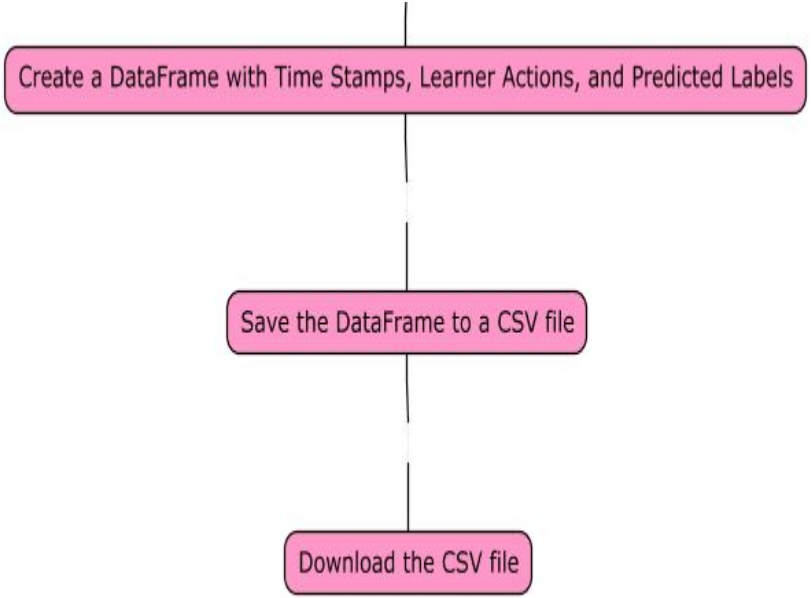
Author: Mayur





Author: Mayur





Author: Mayur

Observation for “sample video”:

1. Frames were extracted at 1 fps.

2. Correctly predicts all Learner Actions (chat gpt, jupyter, excel, other)
along with the features (Jupyter success/error, ChatGPT reading/writing/generating).

3. Model gives prediction results with time stamps at 1 second intervals.

4. Link of Output:
https://docs.google.com/spreadsheets/d/1Ajs2EIQtNdQnARspao1BtdxzI1CHSpDLkTVg-NF6urw/edit?usp=sharing

LA mayur.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[4]

```
Epoch 9/20
26/26 [==============================] - 11s 405ms/step - loss: 0.2278 - acc: 0.9384 - val_loss: 0.3765 - val_acc: 0.8812
Epoch 10/20
26/26 [==============================] - 10s 377ms/step - loss: 0.2142 - acc: 0.9432 - val_loss: 0.3055 - val_acc: 0.8911
Epoch 11/20
26/26 [==============================] - 10s 386ms/step - loss: 0.1980 - acc: 0.9444 - val_loss: 0.3063 - val_acc: 0.8960
Epoch 12/20
26/26 [==============================] - 10s 399ms/step - loss: 0.2029 - acc: 0.9396 - val_loss: 0.3188 - val_acc: 0.8911
Epoch 13/20
26/26 [==============================] - 9s 356ms/step - loss: 0.1850 - acc: 0.9432 - val_loss: 0.2901 - val_acc: 0.9010
Epoch 14/20
26/26 [==============================] - 10s 398ms/step - loss: 0.1872 - acc: 0.9457 - val_loss: 0.2806 - val_acc: 0.8911
Epoch 15/20
26/26 [==============================] - 10s 385ms/step - loss: 0.1864 - acc: 0.9444 - val_loss: 0.2710 - val_acc: 0.9059
Epoch 16/20
26/26 [==============================] - 9s 355ms/step - loss: 0.1783 - acc: 0.9457 - val_loss: 0.2782 - val_acc: 0.8911
Epoch 17/20
26/26 [==============================] - 10s 402ms/step - loss: 0.1673 - acc: 0.9420 - val_loss: 0.2950 - val_acc: 0.8960
Epoch 18/20
26/26 [==============================] - 10s 363ms/step - loss: 0.1706 - acc: 0.9517 - val_loss: 0.2728 - val_acc: 0.8960
Epoch 19/20
26/26 [==============================] - 11s 417ms/step - loss: 0.1634 - acc: 0.9493 - val_loss: 0.2927 - val_acc: 0.8812
Epoch 20/20
26/26 [==============================] - 10s 378ms/step - loss: 0.1627 - acc: 0.9469 - val_loss: 0.2584 - val_acc: 0.9059
7/7 [==============================] - 2s 306ms/step - loss: 0.2584 - acc: 0.9059
Final loss: 0.26
Final accuracy: 90.59%
```

Model Accuracy (with 20 epochs)=90.26%
Final Loss= 0.26

Author: Mayur

Observation for “v2”:

1. Frames were Extracted at 5 fps.
2. Model gives prediction results with time stamps at 0.2 second intervals.
3. Model is inaccurate for first 25 seconds because:
   The user is using anaconda navigator for first 25 seconds and the model has not been trained for anaconda navigator.
4. Link for Output:
   https://docs.google.com/spreadsheets/d/1WTxxY1hqxq8Th-KgVHyCEBD-Q3azJFbH-RcigiEhNiY/edit?usp=sharing

v2_model_predictions_results_updated

File Edit View Insert Format Data Tools Extensions Help

Image	Learner_Actions	Predicted_Label	Time (Seconds)
1	Chat GPT	Gpt Reading	0.2
2	Chat GPT	Gpt Reading	0.4
3	Chat GPT	Gpt Reading	0.6
4	Chat GPT	Gpt Reading	0.8
5	Jupyter	Jupyter Running	1
6	Chat GPT	Gpt Reading	1.2
7	Jupyter	Jupyter Running	1.4
8	Unknown	Other	1.6
9	Unknown	Other	1.8
10	Unknown	Other	2
11	Unknown	Other	2.2
12	Unknown	Other	2.4
13	Jupyter	Jupyter Running	2.6
14	Jupyter	Jupyter Running	2.8
15	Unknown	Other	3
16	Unknown	Other	3.2
17	Chat GPT	Gpt Reading	3.4
18	Unknown	Other	3.6
19	Jupyter	Jupyter Running	3.8

Author: Mayur

Observation for “v1”:

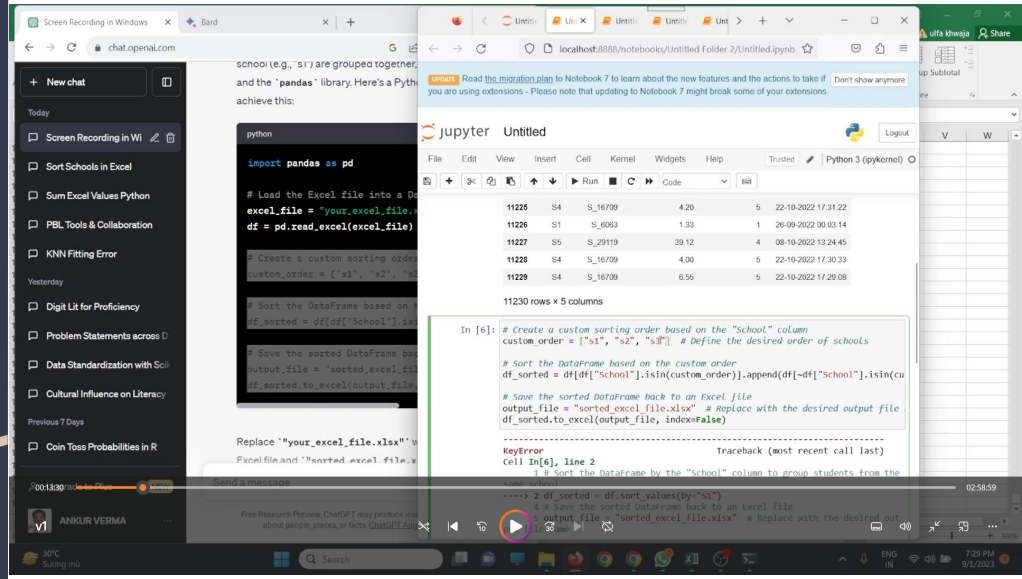
1. Frames were extracted at 5 fps.

2. Discrepancies in Data:

a)Model predictions are inaccurate because the user is operating on multiple tools simultaneously throughout the video.

b) the user is also switching between chatGPT and Google Bard.

3. The model cannot be applied on this video as the user is using multiple simultaneously and the model will fail to classify the environment.



Author: Mayur

Why 1 fps is better than 5fps:

- With frames extracted at 1 fps the video processing requires less infrasture (GPU).
- Model crashes if more than 800 images are given at once for predicting.

- With 1 fps longer video can be processed at once. To process same length of video extracted at 5fps, more RAM is needed, for additional RAM, the user needs to pay.

Author: Mayur

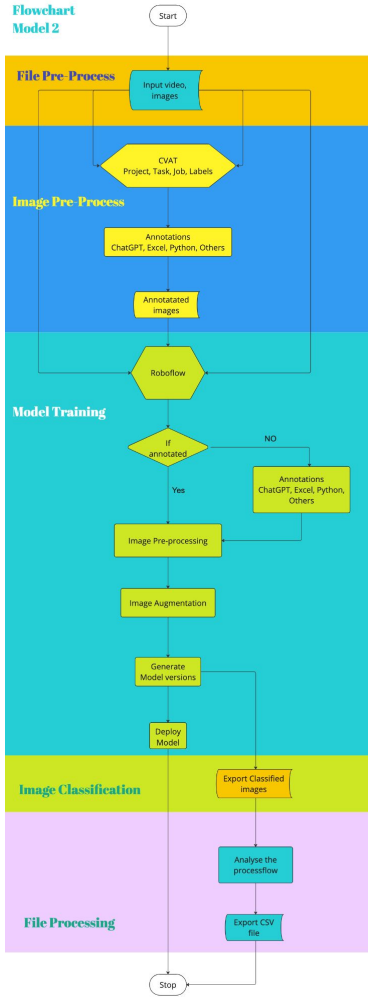
Future Scope:

1. A confusion matrix can be made for each video that is sample video, v1 and v2
   The confusion matrix will help identify which particular feature is being correctly predicted and which one show low accuracy.Based on above analysis more training data can be given for that particular feature.

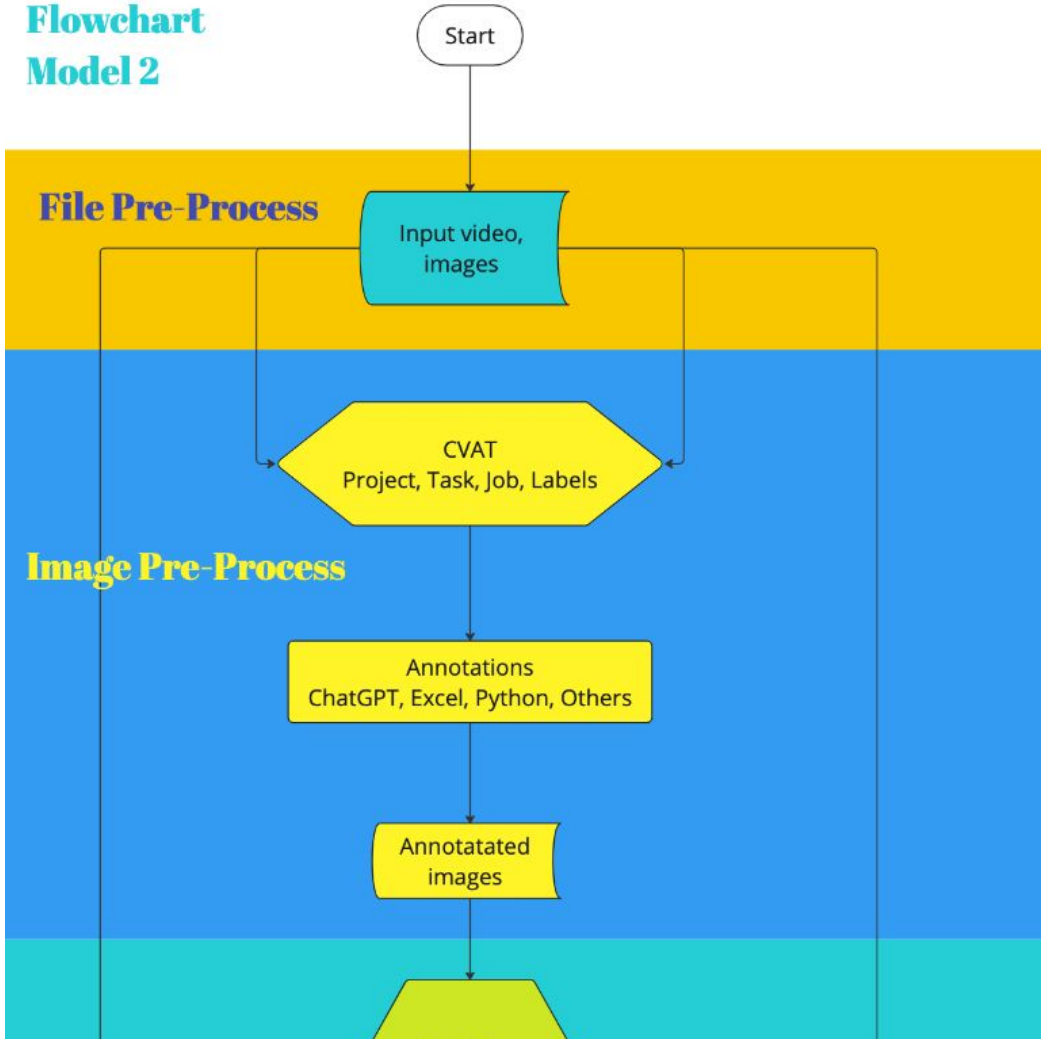
2. Instead of multivariable classification, “one vs. many variables” could be performed so that if the image does not match training data set, it will be put in “Other” Category.

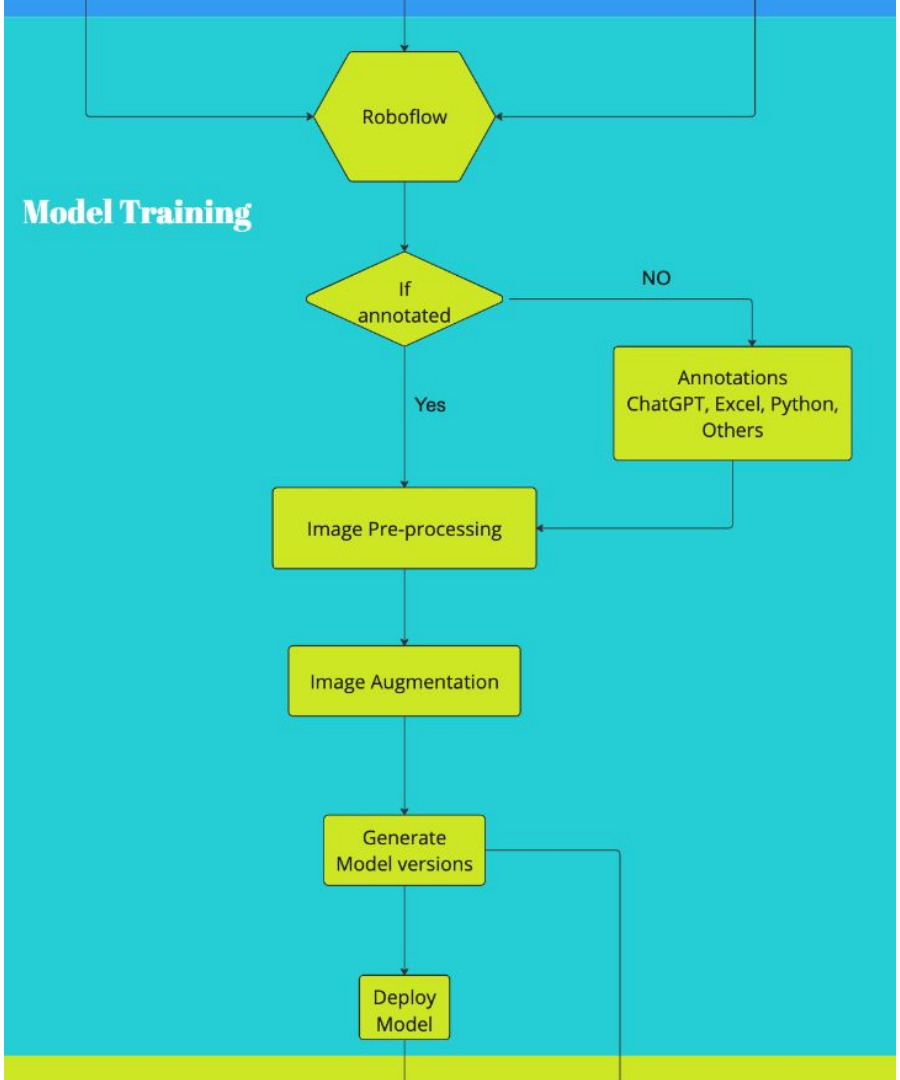
Author: Mayur

MODEL 2


Flowchart
Model 2
Start
File Pre-Process
Input video, images
CVAT
Project, Task, Job, Labels
Image Pre-Process
Annotations
ChatGPT, Excel, Python, Others
Annotatated images
Roboflow
Model Training
If annotated
NO
Yes
Annotations
ChatGPT, Excel, Python, Others
Image Pre-processing
Image Augmentation
Generate Model versions
Deploy Model
Image Classification
Export Classified images
Analyse the processflow
File Processing
Export CSV file
Stop

Flowchart Model 2





Used Pascal VOC (Popular machine learning frameworks and object detection library.)

Image Segmentation

mAP	Precision	Recall
96.0%	95.7%	94.3%

Image Classification

Validation Accuracy

99.1%

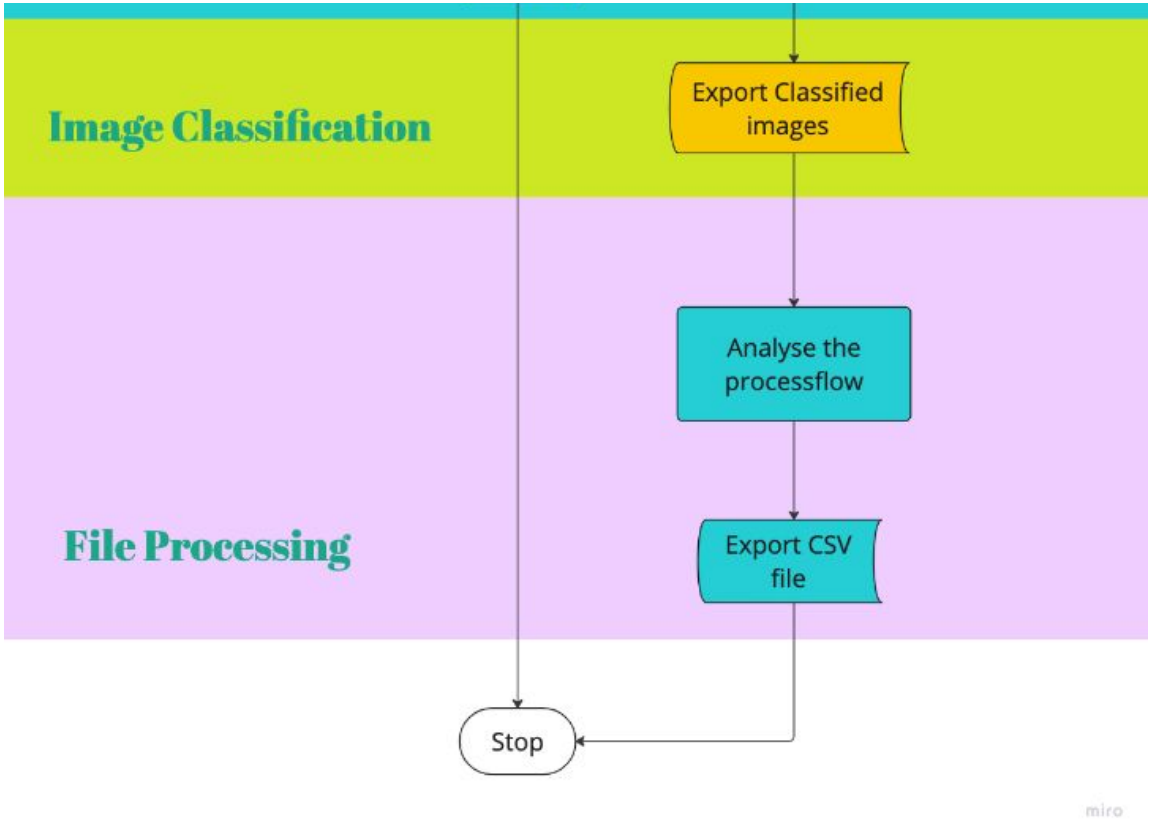
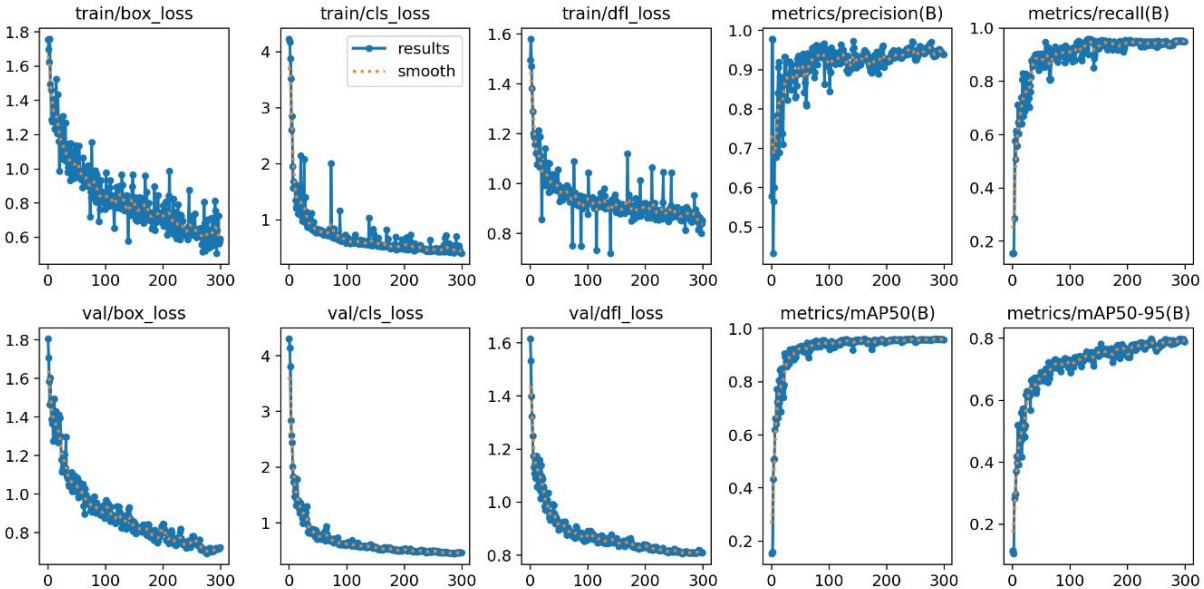


Image Classification
Export Classified images
Analyse the processflow
File Processing
Export CSV file
Stop
miro

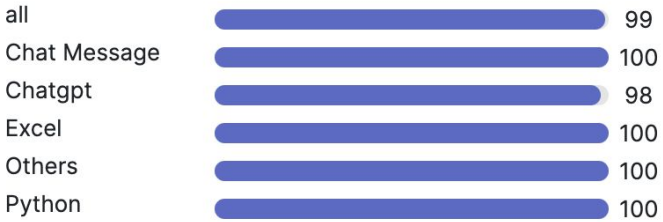
mAP	Precision	Recall
96.0%	95.7%	94.3%

Validation Set | Test Set | Training Graphs



Validation Set | Test Set | Training Graphs

Average Precision by Class



mAP is equal to the average of the Average Precision metric across all classes in a model.

Precision measures how often your model's predictions are correct.

Recall measures what percentage of relevant labels were successfully identified.

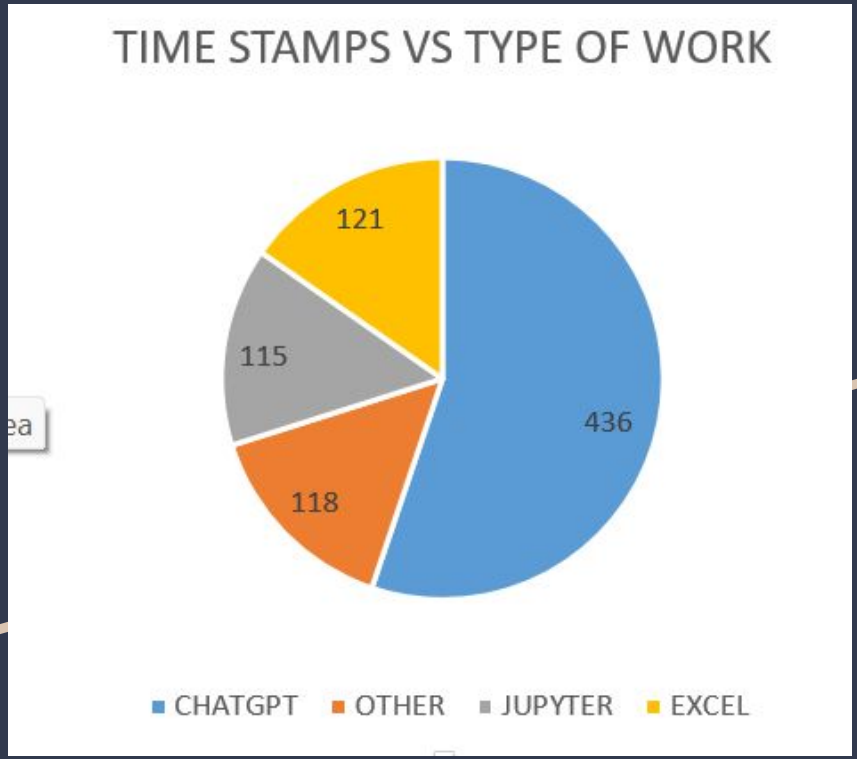
OBSERVED DATA AFTER RUNNING THE MODEL

Time Stamp	Learner_A	Predicted_Label
1	Chat GPT	Gpt Reading
2	Chat GPT	Gpt Reading
3	Chat GPT	Gpt Reading
4	Chat GPT	Gpt Reading
5	Chat GPT	Gpt Reading
6	Other	Other
7	Other	Other
8	Other	Other
9	Other	Other
10	Other	Other
11	Other	Other
12	Jupyter	Jupyter Running
13	Jupyter	Jupyter Running

4 Categories of DATA that our team generated from the model

1. Excel
2. Chat GPT
   - GPT Writing
   - GPT Generating
   - GPT Reading
3. Jupyter Notebook
   - Jupyter Success
   - Jupyter Running
   - Jupyter Error
4. Others

Number of Frames vs Type of Work Done



1.The model predicted 800 images.

2. Model accurately predicts all learner actions except “Jupyter Running”.

3.Accuracy: 91% Overall Accuracy of the trained model on the validation set.

FUTURE SCOPE

- We will try to predict ‘Jupyter Running’ learner action in future using Yolo.
- Using Yolo to extract data into ‘csv’ format is also identified for future.

THANK YOU

Feedback/Questions ?