UE18CS390B – Capstone Project Phase – 2

SEMESTER - VII

END SEMESTER ASSESSMENT

Project Title : Visual Question Answering On Statistical Plots

Project ID : PW22MHR02

Project Guide : Dr.Mamatha H.R

Project Team : Sneha Jayaraman PES1201802825
Sooryanath I T PES1201802827
Himanshu Jain PES1201802828

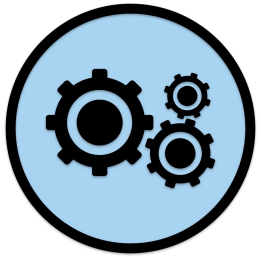
OUTLINE

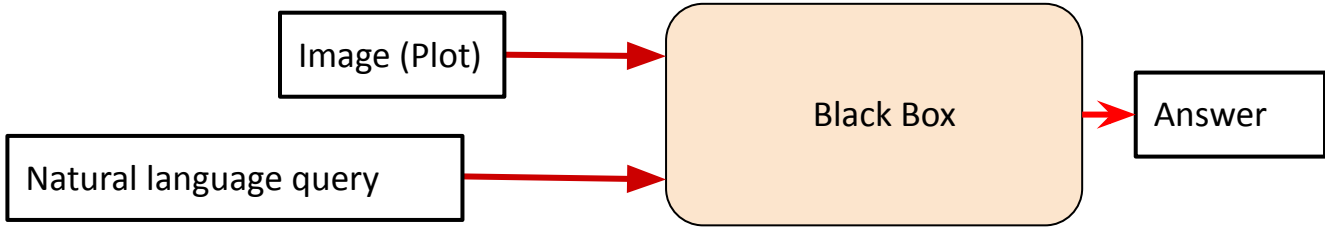
- Abstract
- Team Roles and Responsibilities.
- Summary of Requirements and Design (Capstone Phase - 1)
- Summary of Methodology / Approach (Capstone Phase - 1)
- Design Description
- Modules and Implementation Details
- Project Demonstration and Walkthrough
- Test Plan and Strategy
- Results and Discussion
- Lessons Learnt
- Conclusion and Future Work
- References

ABSTRACT

Visual Question Answering system in general, generates an answer to queries posed in natural language with the help of information extracted from the input image. Statistical charts are an easy way to generalize and understand the data. **Incorporating the ability to answer the questions on statistical charts confined to our scope is the aim of this research**. In the development of our model, questions must be related to answers from fixed vocabulary or answers that can be extracted from the bounding box representation of an image or answers that can be queried from a structured table generated using visual elements. **Plots related to bar plot, line plot, and dot plots have been considered.**

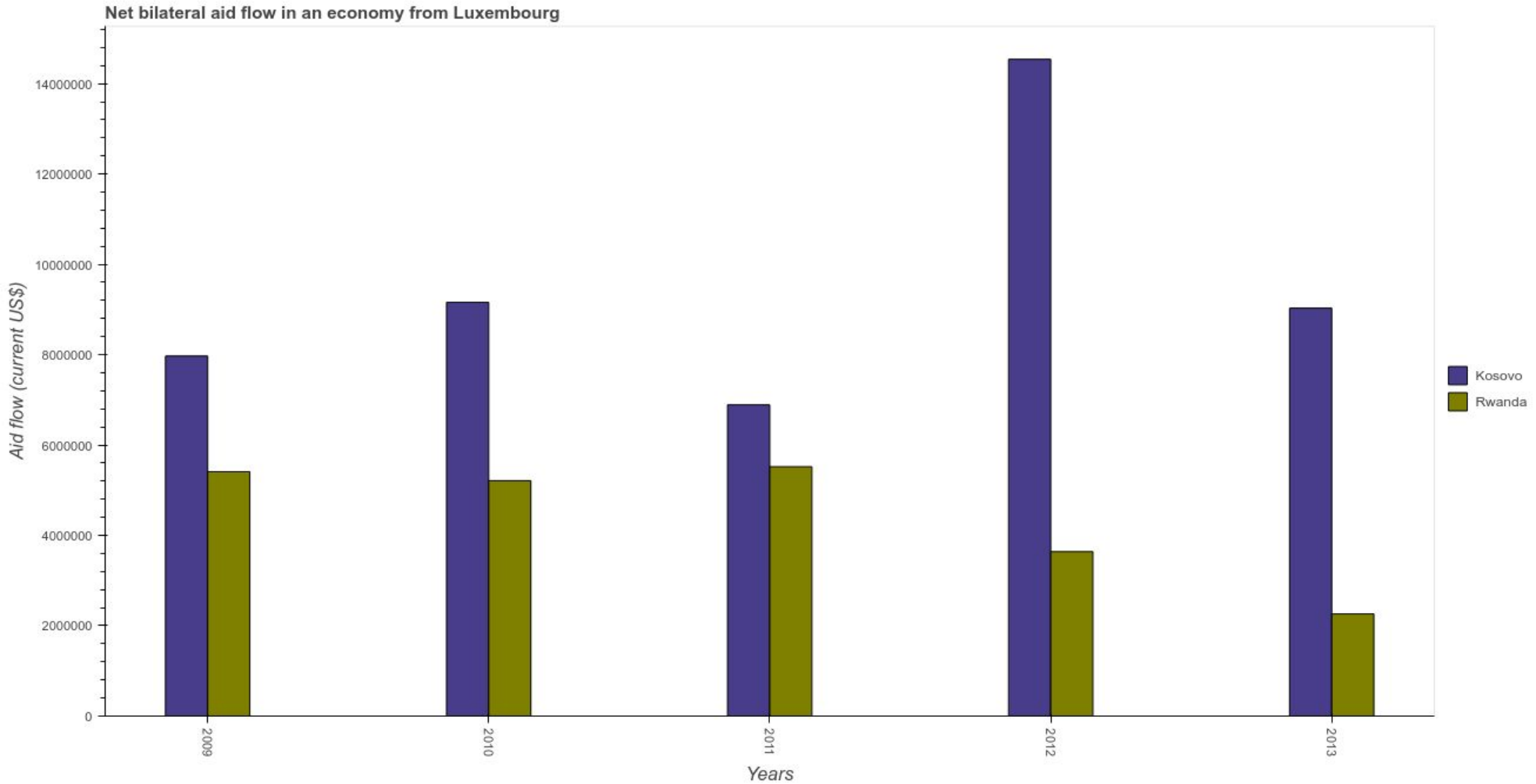
Domain: Deep Learning, Object Detection & NLP





SCOPE

Scope: We will deal with a certain section of **Plots = Dot , Line , Bar[Hbar , Vbar , Grouped]** and **Questions = Open-ended, Boolean {Support | Refute} .**



TEAM ROLES AND RESPONSIBILITIES

Team Member	Role & Responsibilities
Sneha Jayaraman	• Re-producing the Sempre model, Detectron-1 model presented in the base paper • OCR and Semi-structured Table Generation Stage • TAPAS, and TabFact model • Testing and Evaluation of Tapas and TabFact on line and dot plots.
Sooryanath I T	• Dry run of existing models (PlotQA) • Data Preparation & Preprocessing + Platform Setup • Plot Element Detection Stage [Train+Test]. • Developed a sample Web-App demonstrating E2E functionality.
Himanshu Jain	• Data preparation and Preprocessing • Binary Classifier • Tapas and Tabfact model • Testing and evaluation of Tapas and Tabfact on horizontal and vertical bar plots.
Common Tasks	• Strategic decision making & planning. • Brainstorming over various approaches. • Critiquing the work done by peer. • Documentation of Report, IEEE draft and other deliverables.

REQUIREMENTS AND DESIGN

- In this section we list the major requirements and the High Level Design we intended to conceive.

Major Functional Requirements

- Inputs : Statistical Plot (within scope) & Query (within scope)
- The blackbox must accept the inputs and process them to extract the bounding box annotations and process them in a suitable manner to provide answer to the query posed.
- Output : A numerical floating point / integer or Textual or Boolean answer.

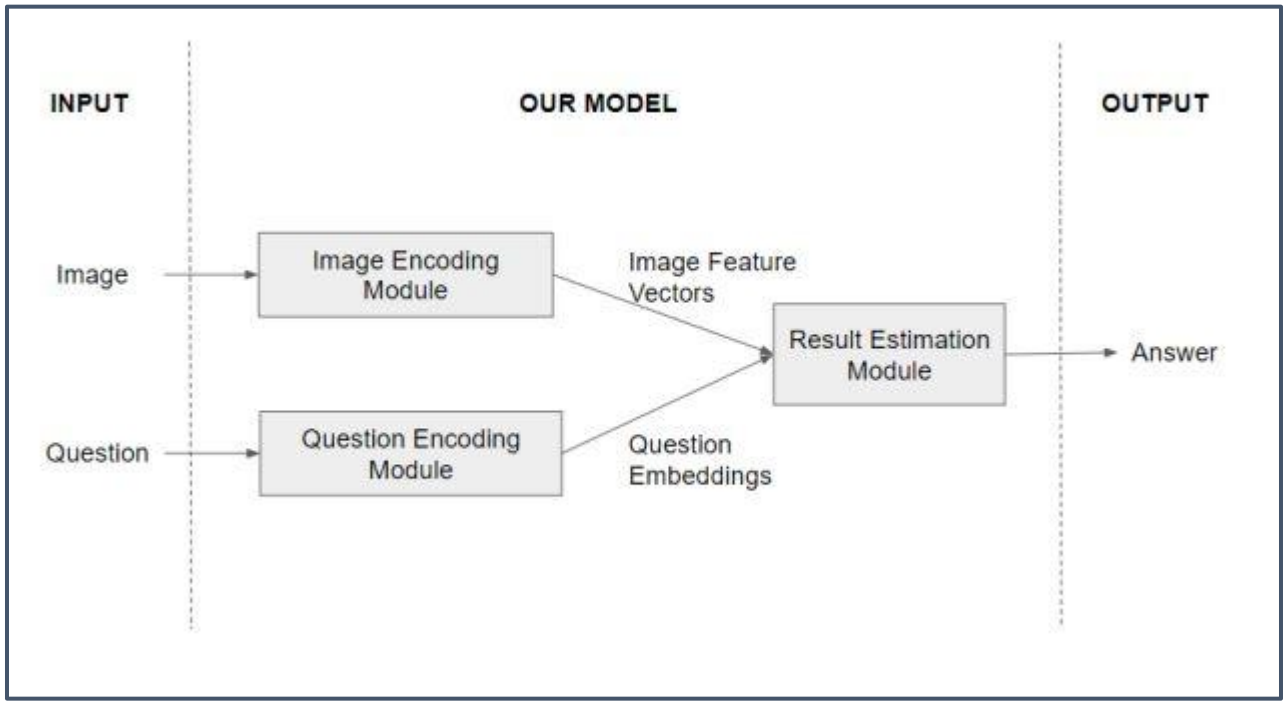
REQUIREMENTS AND DESIGN

Major Non Functional Requirement

- Usability :
  - The trained model must be available at ease to use it just by choosing an image input from the local drive or file-system .
  - Similar to drag and drop or attaching files.
- Dependency :
  - The stages in the pipeline must interface well with the upstream and downstream neighbors

REQUIREMENTS AND DESIGN

➢ ***High Level Design (As conceived in Phase - 1)***



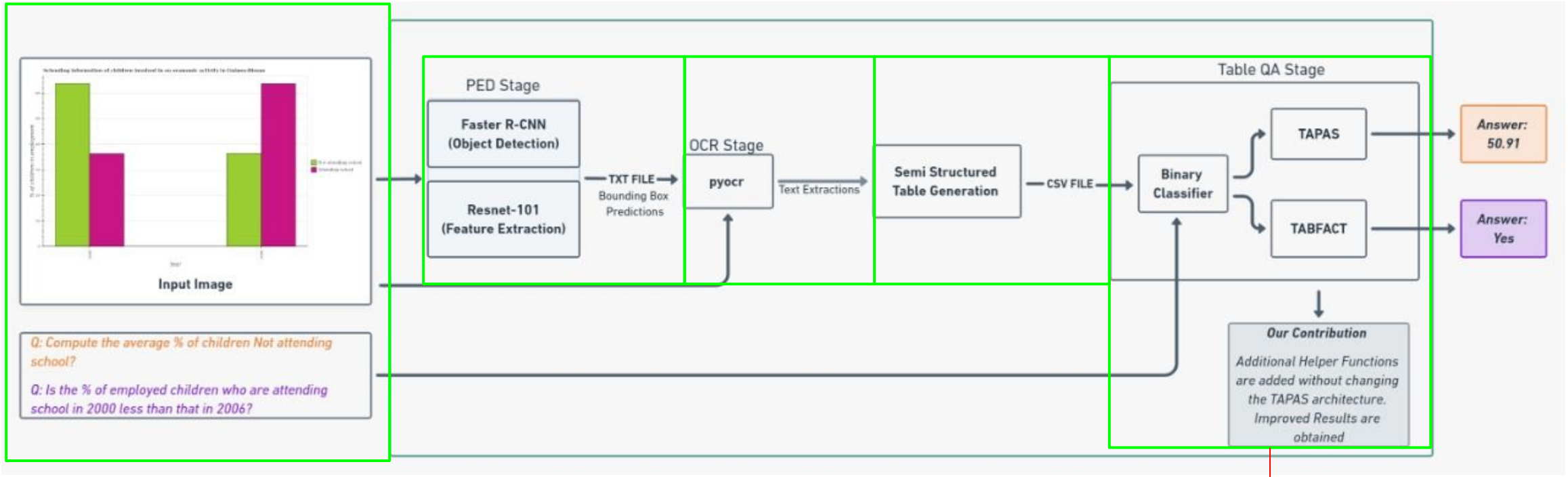
The Dataset Split-Up

- The 2 Major datasets related to this problem statement are **DVQA** & **PLOT_QA** datasets.
- However DVQA contains graphs of the bar category only.
- Whereas the PLOT_QA dataset contains plots of type {vbar , hbar , dot , line} ,hence we chose this dataset.
- The distribution of Plot_QA dataset is displayed below

Dataset Split	#Images	#QA pairs
Train	157,070	20,249,479
Validation	33,650	4,360,648
Test	33,657	4,342,514
Total	224,377	28,952,641

Category	Count_of_Samples_In_Test_Set
dotline	5574
H_bar	11292
Line	5549
V_bar	11242

OUR APPROACH : *END-TO-END PIPELINE*



- TableQA Stage is where the novelty is induced
- We will detail it out in the upcoming slides

METHODOLOGY / APPROACH

➢ This section discusses the methodology supported by the low level design , a peek into the model (stage-wise), the benefits and drawbacks.

Approach: The design consists of 4 stages

1) **Plot Element Detection Stage**
2) **Optical Character Recognition Stage**
3) **Semi-Structured Table Generation Stage**
4) **Table Question Answering Stage**
   - **TAPAS** model to answer open-ended questions
   - **TABFACT** to answer Yes/No questions

Advantages of Our Approach

- The need for 2 separate pipelines to handle boolean and descriptive questions is eradicated.

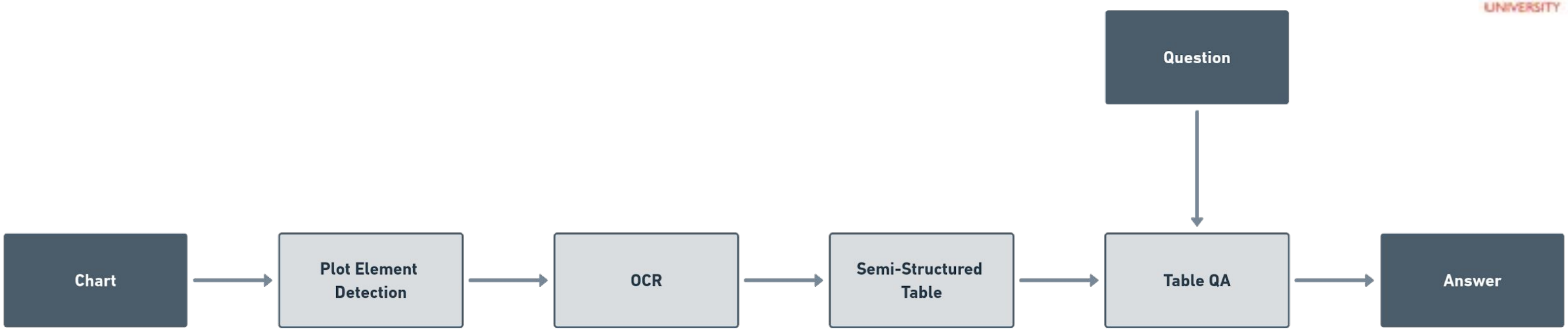
- Both the boolean and in-vocabulary queries can be handled within a single pipeline , due the presence of binary classifier & the new approach for Table-QA Stage.

- Generates a tabulated format of plot data , which can be made available to the user if needed.

Drawbacks of Our Approach

- Limited to a few types of plots as defined in the scope
- Limited by the types of questions it can answer
- Scope of improvement to match human Accuracy

MODULES AND IMPLEMENTATION DETAILS

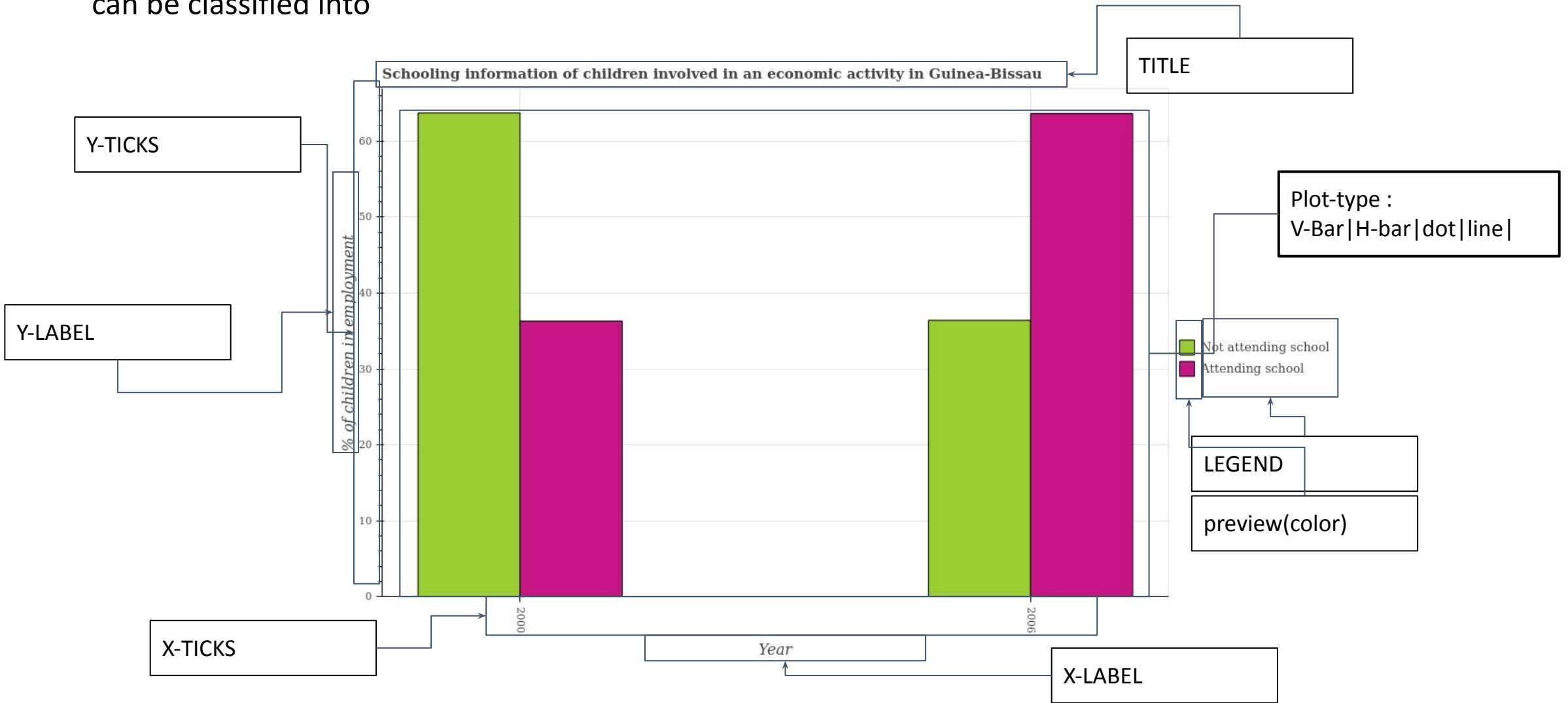


- As indicated before , the model pipeline consists of 4 phases/stages .
- In the following section we shall elaborate the minute details and the functional aspects of the stages
- The specifications , performance and dependency will be discussed as well.

STAGE : 1

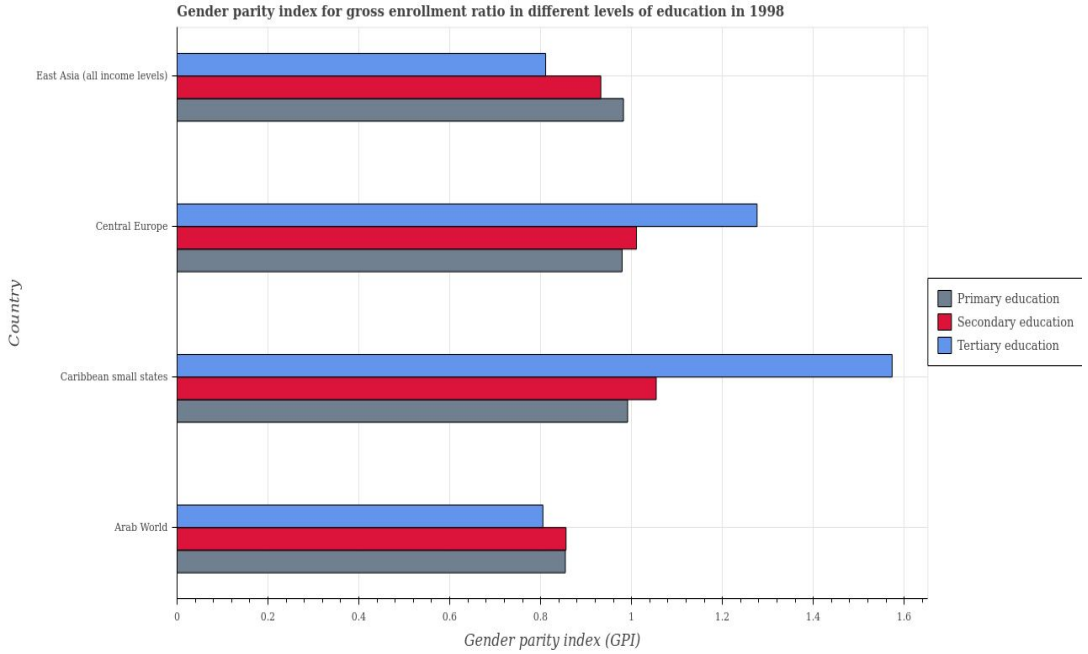
Plot Element Detection

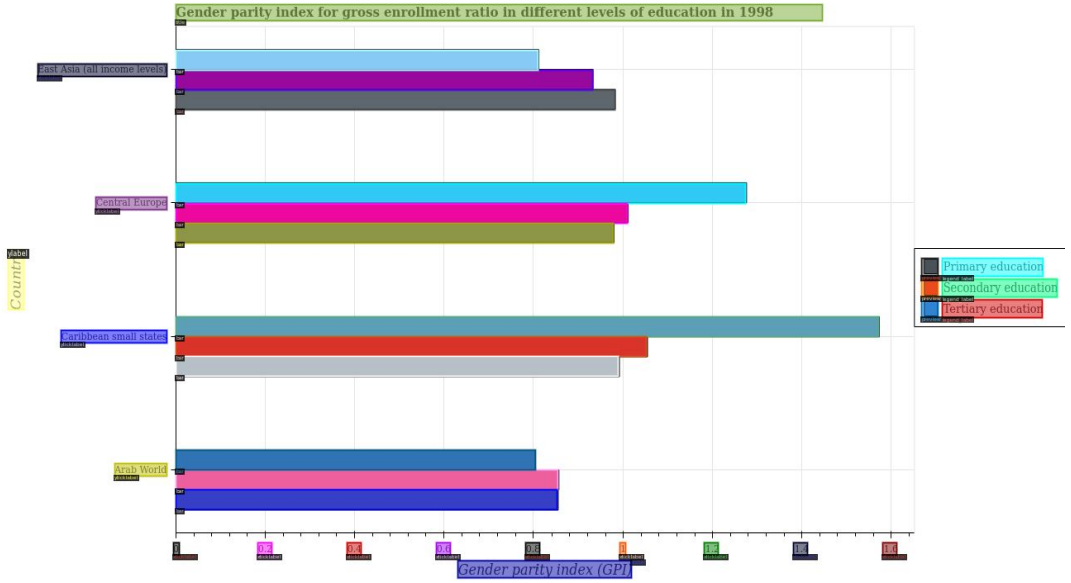
- There are a total of 10 object {here : plot_elements } classes that every image {here : plot_image} can be classified into



Plot Element Detection

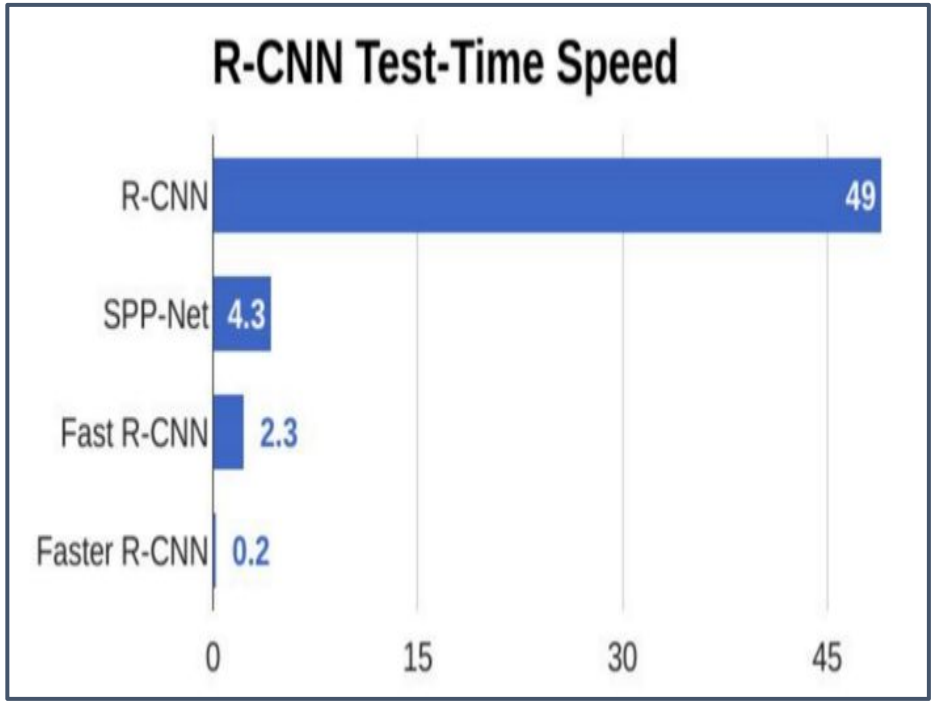
- From previous slides we noticed that there are several objects (plot elements) inside an image(the graph).
- While an image classification network can tell whether an image contains a certain object or not, it won't say where in the image the object is located. We are in the quest of **locating the plot elements** and **not just classifying the plot.** Hence we just don't need just a CNN ✗ we need an object detection neural network like R-CNNs ✓

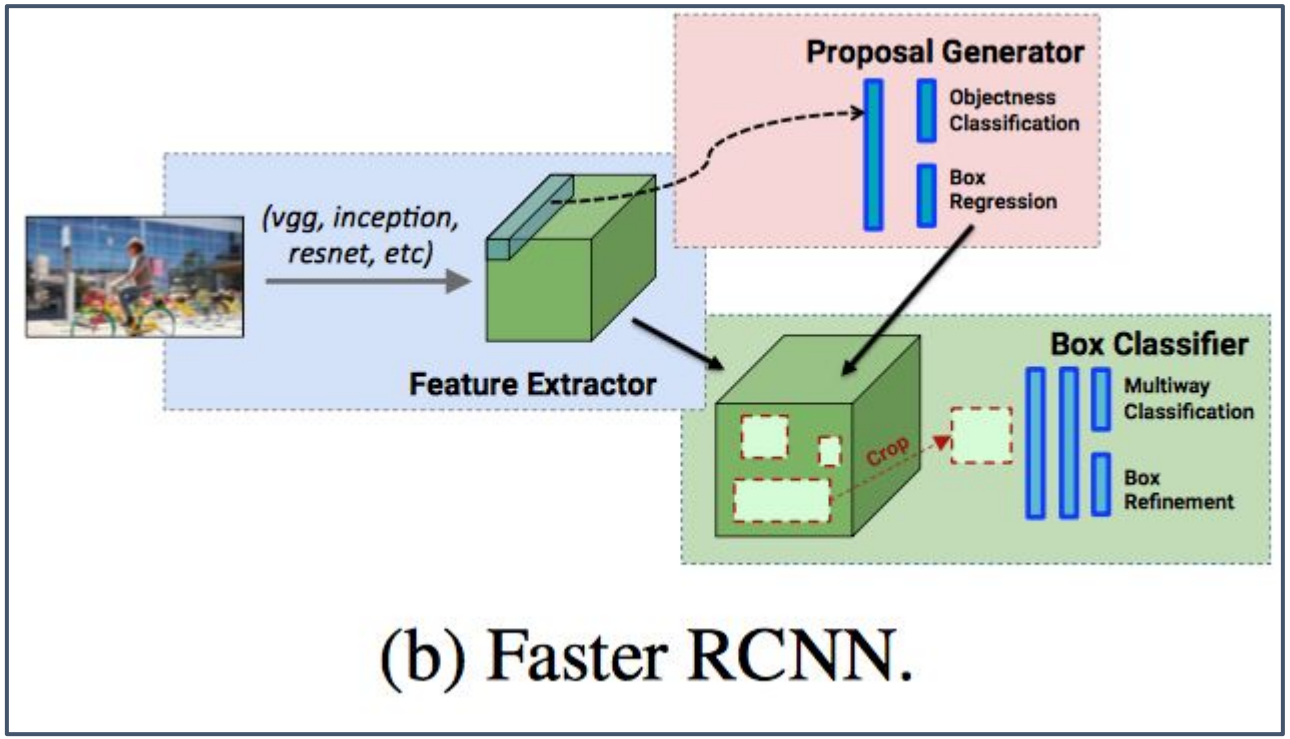




Plot Element Detection

- The dataset is **correlated and registered for training in the detectron-2 tool** , we need to decide over the **object detection model** .
- We chose **Faster-RCNN** as our object detection model which is a **descendant from the R-CNNs series**.
- The heuristic behind doing so can be inferred from the below observations and also due to the presence of the RPN (Regional proposal network is known for locating feature targets accurately) .

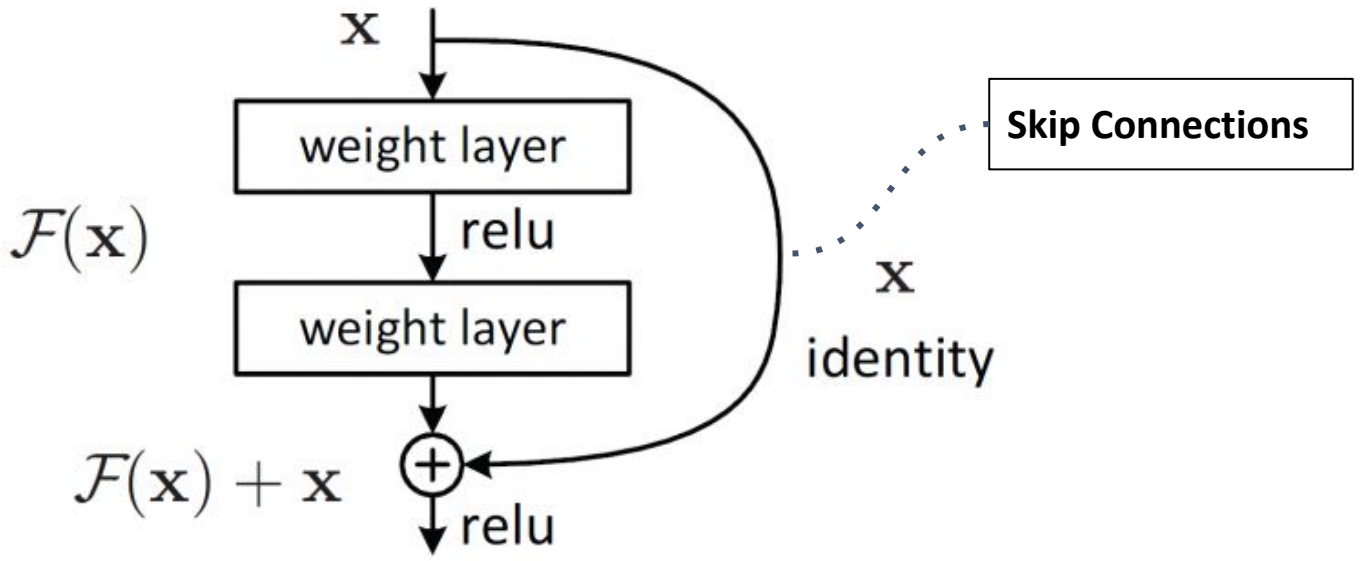




(b) Faster RCNN.

Plot Element Detection

- We also need to decide on the **feature extractor model** for serving as the **backbone/feature-extractor** for our faster-RCNN Setup.
- We Chose **Resnet-101** to be our feature extractor due to **skip connections** and **residual blocks** which can reduce the problem of vanishing gradients in a very deep network like Resnet-101.



Plot Element Detection

- Now that we have decided our model combinations as per the proven - observations ,we head on to training.
- The training environment specification is as follows
  - **Platform** : Google Colab Pro
  - **RAM** : 26GB
  - **Disk** : 110GB
  - **GPU** : 16GB , Tesla - T4
  - **Training Data Size** : 6.x GB
- We implemented the data-processing + prep and the training using **Detectron-2** an open-source research tool for deep learning.

Why did we choose Detectron-2?

- **Detectron-2** is faster , flexible and vast in terms of configuration,models and implementation due to its API availability when compared with its parent.
- We also generated our custom YAML file for future training.

STAGE : 2 -3

OCR and Semi-structured table generation stage

- Given the json file for the plot image containing bounding box values (this comes from the previous stage), we would want to format the data into a txt file.
- Next, we would want to extract the plot information into a semi-structured table from which we can answer questions.
- This is what the OCR and Semi-structured table generation stage does.

OCR Stage

There are a total of 10 classes that every plot element can be classified into:

TEXTUAL ELEMENTS

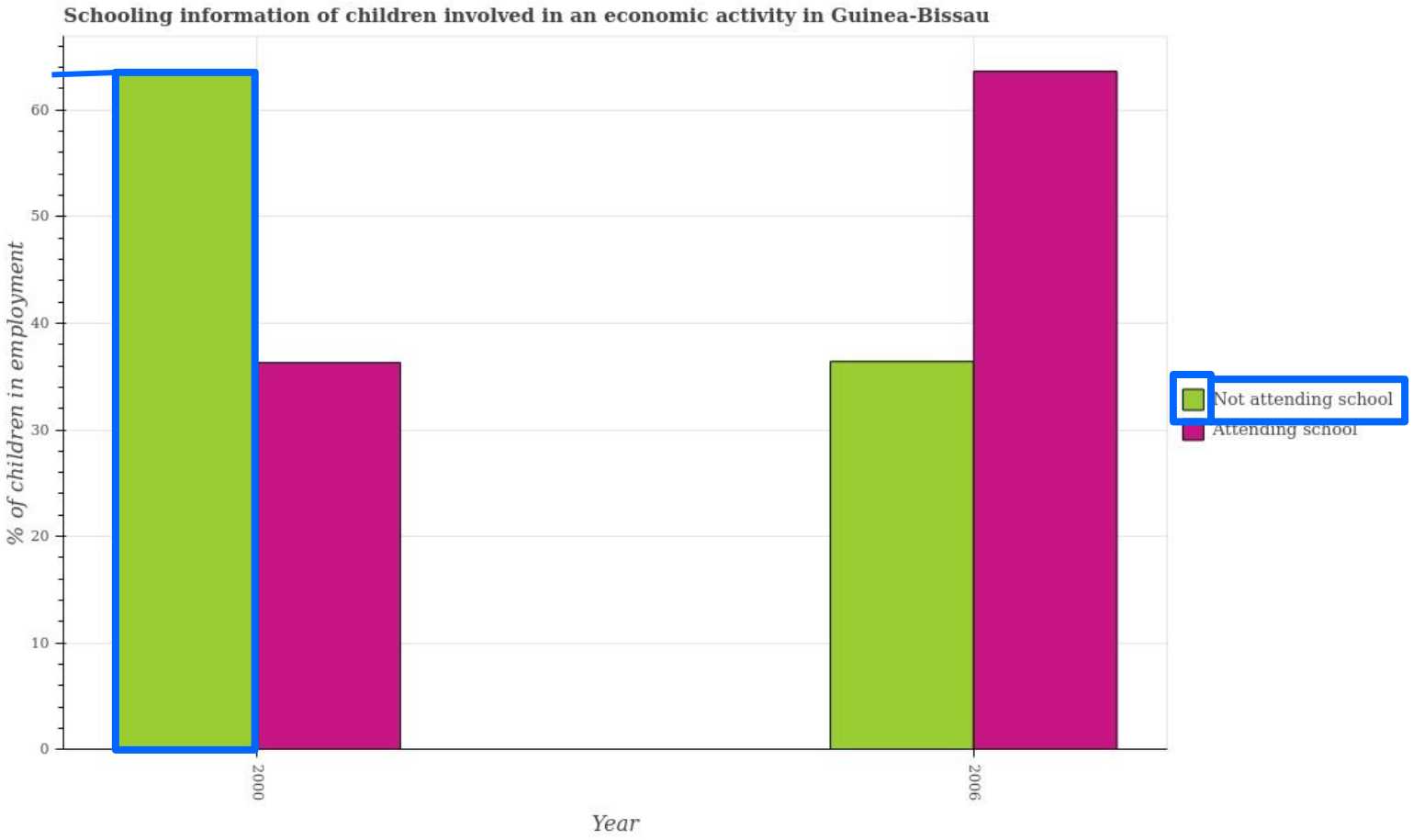
- title
- xlabel
- ylabel
- xticklabel
- yticklabel
- legend label

VISUAL ELEMENTS

- bar
- line
- dot line
- preview - legend color

- The OCR module used is pyocr which is a wrapper for the Tesseract OCR engine.
- Textual data is extracted from all textual elements using the OCR module.

Semi-Structured Table Generation Stage



Semi-Structured Table Generation Stage

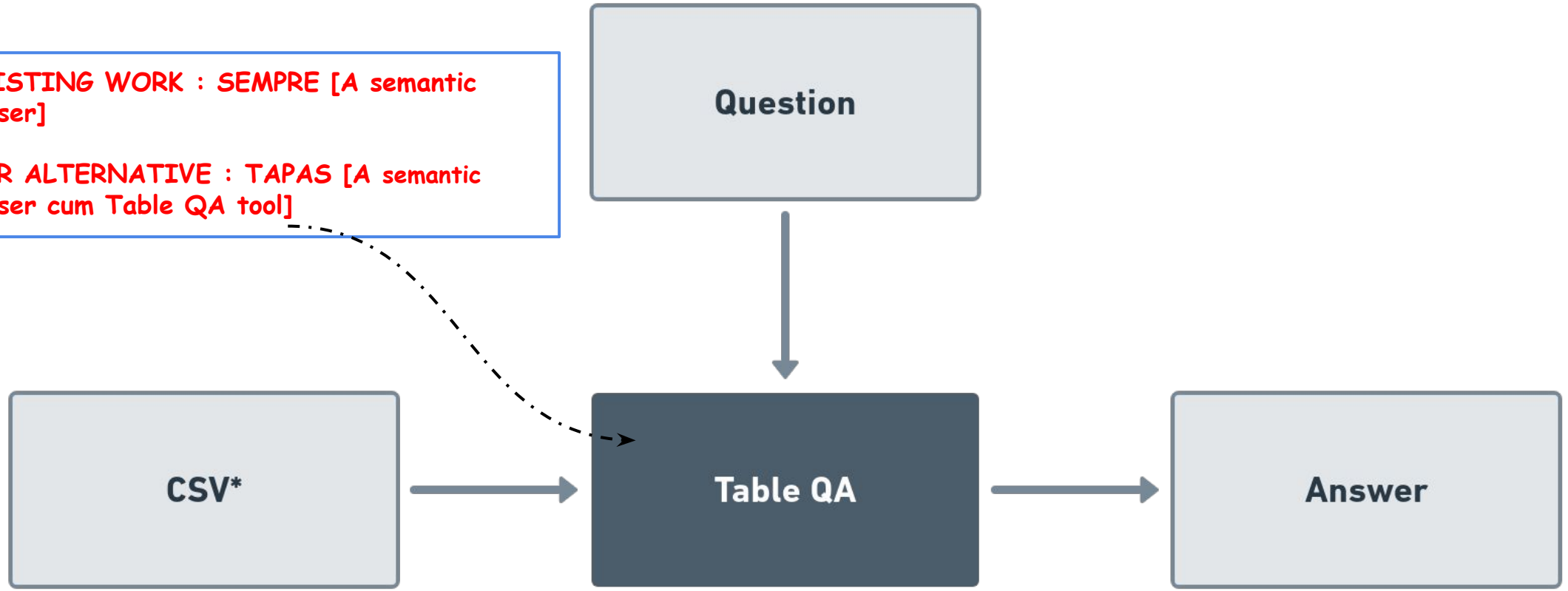
- This stage is responsible for mapping legend values to the legend color, x-ticks to the x-axis label and the y-ticks to the y-axis label.
- This is done by associating the legend / x-tick / y-tick value bounding box to the closest legend color / x-axis / y-axis boundary respectively.
- For the visual elements, each element is associated with an axis, and a corresponding legend. The color of the visual element is matched with the legend colors, and the legend of the closest match is associated with the element.
- To find the value associated with the bar, the information of height is taken from the bounding box representation, and the closest y-tick is mapped.
- Doing this for all visual elements will fill the table and result in a table that is stored as a comma separated file.

STAGE : 4

Table-QA Stage

EXISTING WORK : SEMPRE [A semantic parser]

OUR ALTERNATIVE : TAPAS [A semantic parser cum Table QA tool]



* Generated from the pipeline

Table-QA Stage

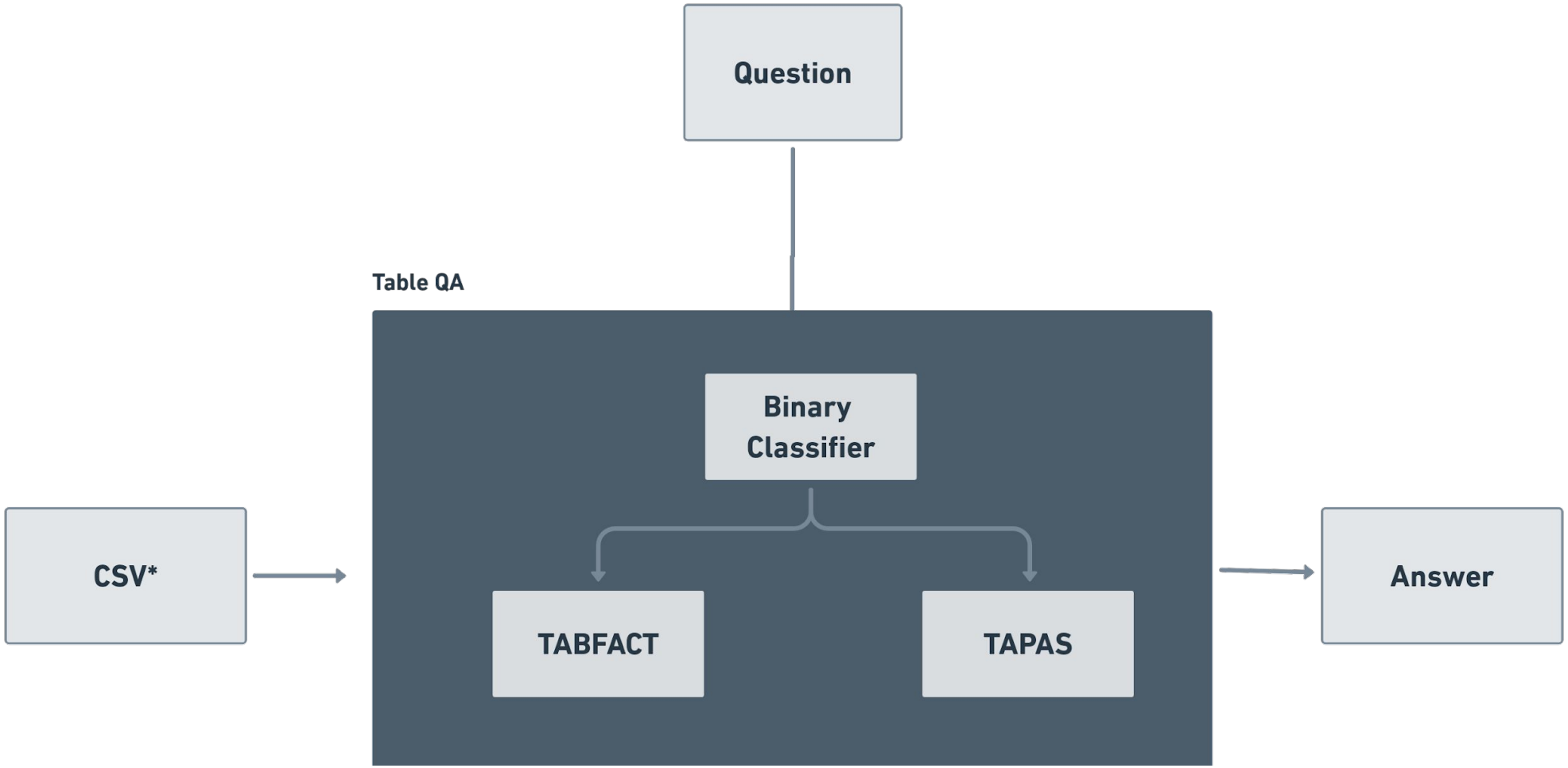


Table-QA Stage

- The prerequisite to this stage is a table that contains all the data from the plot.
- We have made use of ***Google's TAPAS*** to answer questions from this tabular data.
- Tapas selects a subset of table cells as deemed appropriate to the question posed and applies aggregation/retrieval operations on top of them.
- It extends BERT architecture with additional embeddings that captures the tabular structure, with two additional classification layers for selecting cells and predicting a corresponding aggregation operator.

Table-QA Stage

- TAPAS provides us with pre-trained models trained on different datasets for end tasks.

- We have made use of two such models for our purpose.

- To be able to answer additional variety of questions such as *range based operations , ratios , differences etc*, we have added our custom operations supported with our custom helper functions to suit the desired output.

Table-QA Stage

- The Usage of **Tapas** in the Table-QA stage is what differentiates us from the base paper [Plot-QA]
- The heuristic behind doing so is discussed next.

Table-QA Stage

Comparison between TAPAS and Sempre

Works For Questions on	TAPAS	Sempre
Count	Yes	Yes
Sum	Yes	Yes
Average	Yes	Yes
Max and Min	Yes	Yes
Selection operation on cell (Select and Project Ops)	Yes	Partial
Selection operation on column	Yes	No
Aggregation on subset of cells	Yes	No

Table QA-Stage

- Given a question, it can be classified into one of two types.
  - The *first type* corresponds to **open-ended questions** that have an unrestricted answer domain.
  - The *second type* corresponds to questions that require a **Yes/No** (binary) answer.

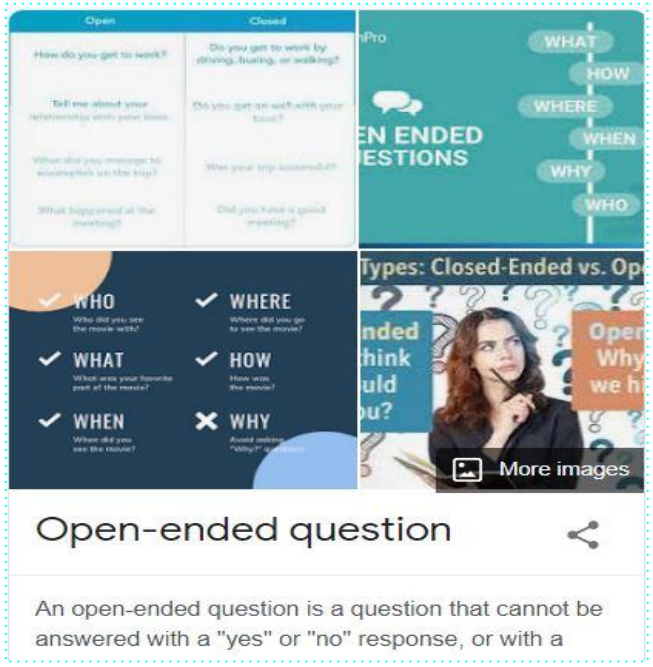


Table QA-Stage

OPEN ENDED QUESTIONS

To handle open-ended questions, we have made use of the existing TaPas (Table Parsing) model that has been trained on the **WikiTables Questions dataset** with intermediate pre-training.

- This model can handle 3 types of aggregation operations - SUM, COUNT, AVERAGE + Simple data - retrieval
- To add to the capabilities of this model, we have added other operations such as: RATIO, DIFFERENCE, MEDIAN, TREND, RANGE and QUARTILES

Table QA-Stage

YES/NO QUESTIONS

To handle questions that require a Yes/No answer, we have used a TaPas model trained on the **TabFact dataset**. This is a dataset used for table entailment and fact verification.

- We have extended its capabilities by adding other operations like mentioned in the previous slide (QA-Types)

Table-QA Stage

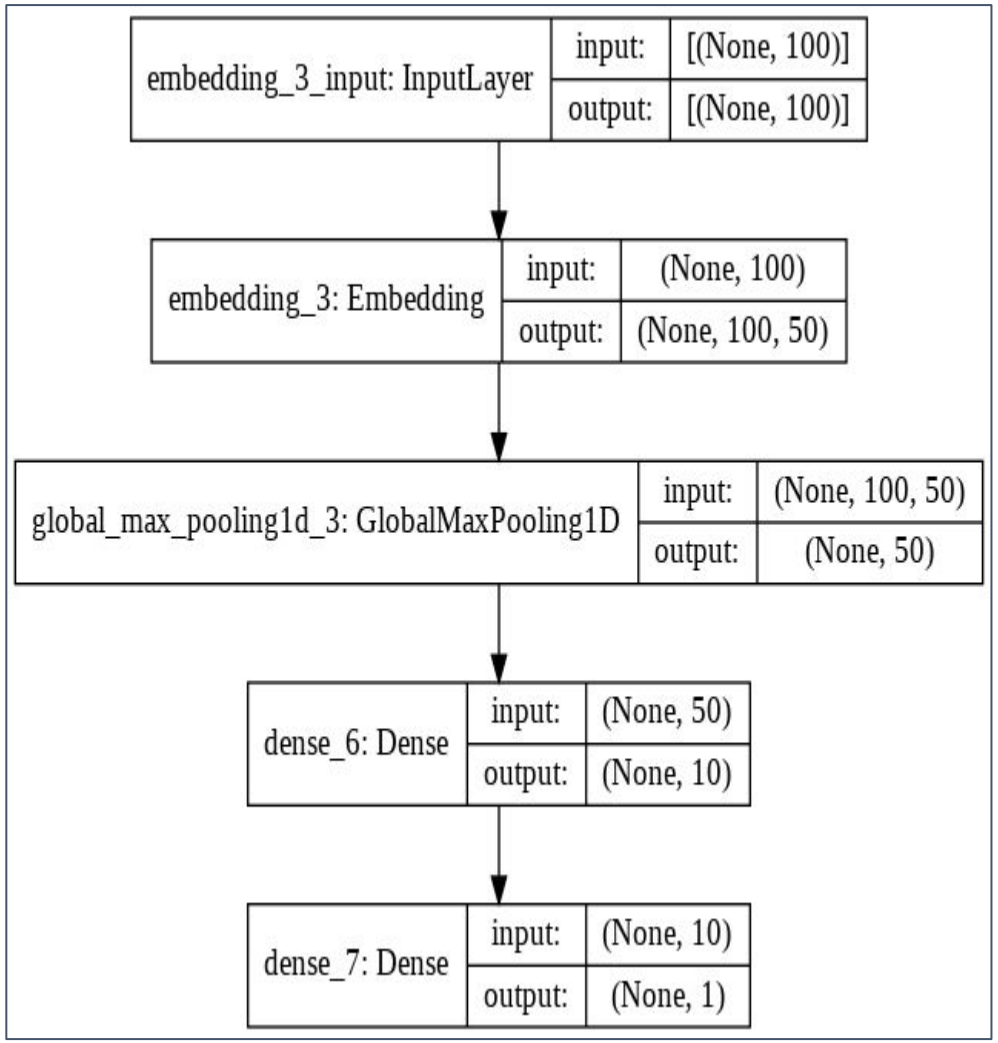
Type of queries we deal with.

Slice	Words
Aggregations	total, count, average, sum, amount, there, only
Superlatives	first, highest, best, newest, most, greatest, latest, biggest and their opposites
Comparatives	than, less, more, better, worse, higher, lower, shorter, same
Negations	not, any, none, no, never

Table QA-Stage - Question Classifier

- An *important aspect* here is that given an input question we would need to know what type of question it is (whether it is an *open-ended question* or a *yes/no* question).
- For this, we have implemented a **binary question classifier.**

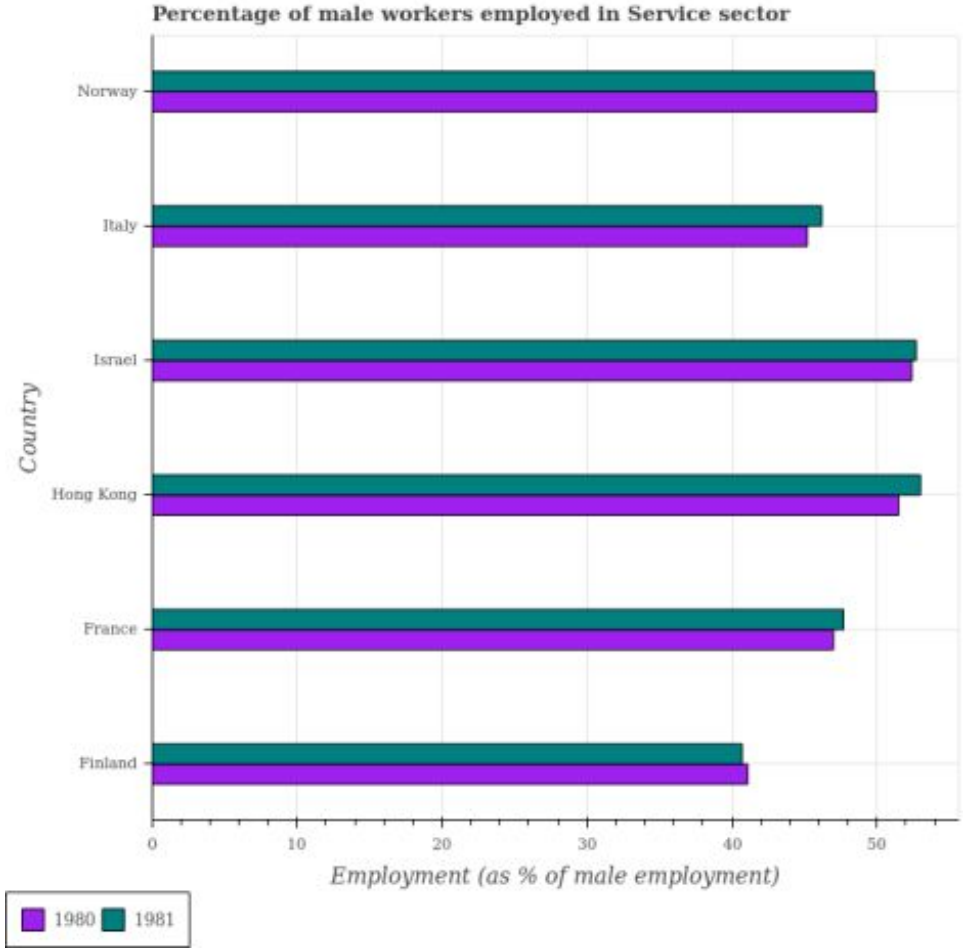
The question classifier is a deep-learning model trained on the PlotQA Questionnaire dataset , so that the questions of different types {open-ended/Yes-No} can be routed to the appropriate methods to table question answering.



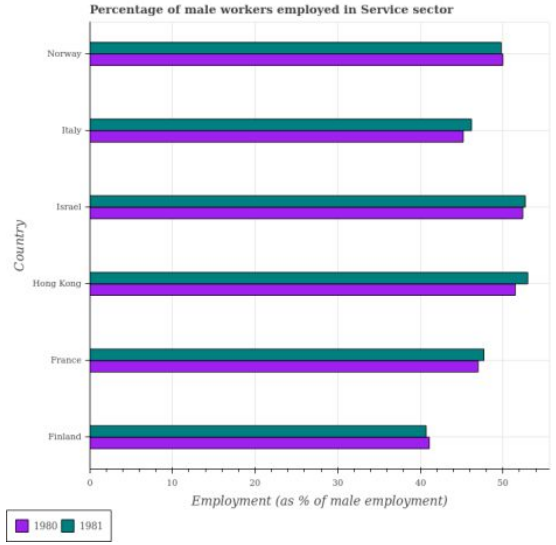
Architecture Of The Binary Classifier

TYPES OF QUESTIONS ADDRESSED

EX: INPUT IMAGE



TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

1. Count

Q: what are the total number of countries ?
A: 6

2. Sum

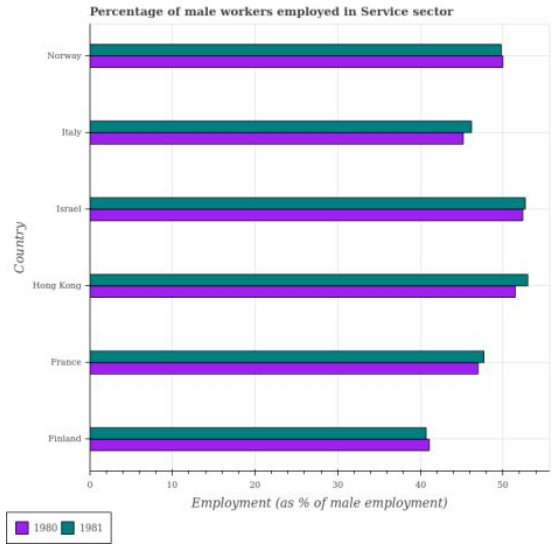
Q: what is the total number of male workers employed in 1980 ?
A: 296.2171903097152

3. Average

Q: what is the average number of male workers in the year 1980
A: 49.36953171828586

Q: what is the average number of male workers in the year 1981
A: 49.26827598218275

TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

4. Minimum

Q: across all countries, what is the minimum percentage of male workers employed in service sector in 1980
A: 41.56915064054187

5. Maximum

Q: across all countries, what is the maximum percentage of male workers employed in service sector in 1980
A: 54.191569798758344

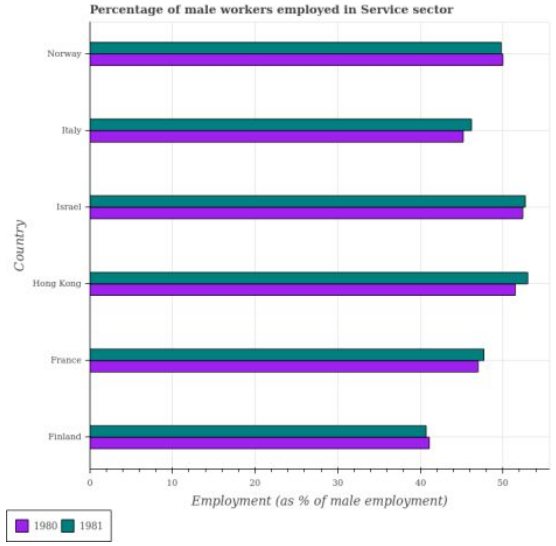
6. Difference

Q: what is the difference between the average number of male workers employed for the year 1980 and 1981
A: DIFFERENCE = 0.10125573610311278

Q: what is the difference between the number of male worker employed for the country france in 1980 and 1981
A: DIFFERENCE = 0.5687026318456105

- Keyword = “difference between”
- The two entities must be separated by "and"

TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

7. Median

Q: what is the median number of male workers employed in the year 1980
A: MEDIAN = 50.22946206732409

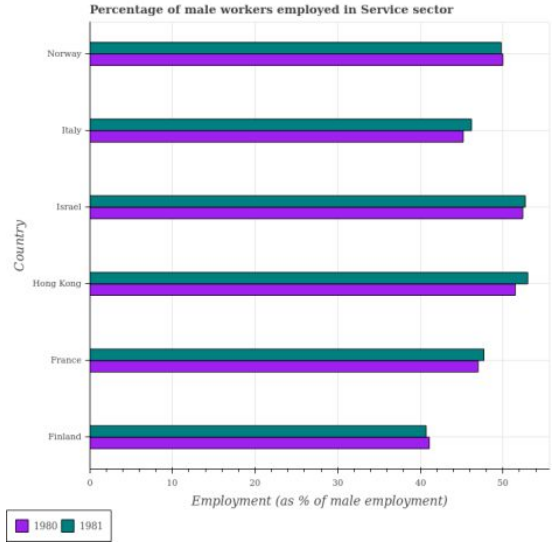
- Keyword = “median“
- Column name for which the median has to be found

8. Ratio

Q: what is the ratio of male workers employed in 1981 to 1980 for the country hong kong A: RATIO = 1.0093722657385955

- Keyword = “Ratio“
- QUANTITY_1 "to" QUANTITY_2

TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

9. Trend (Increasing or Decreasing)

Q: what is the trend of male workers employed for the countries finland, france, hong kong in 1980
A: TREND = INCREASING

Q: what is the trend of male workers employed for the countries israel, italy in 1980 A: TREND = DECREASING

Q: what is the trend of male workers employed for the countries israel, italy, norway in 1980 A: TREND = NONE

- Keyword = “trend“
- List of comma separated entries followed by "in" COL_NAME

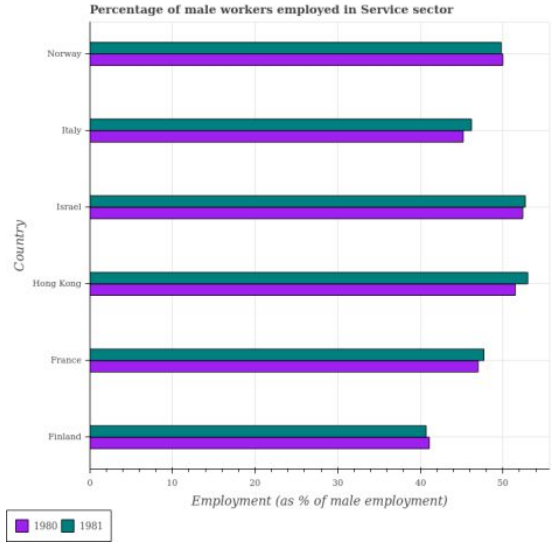
10. Selection operation on cell

Q: what is the number of male workers employed for country france in the year 1981 A: 48.40140328049439

11. Select operation on cell after applying aggregation operation

Q: which country has the minimum number of male workers employed in the year 1981 A: Finland

TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

12. Selection and Aggregation operation on subset of rows

Q: what is the sum of male workers employed for the countries france, finland in the year 1980 A: 90.53925655288188

Q: what is the average number of male workers employed for the countries france, finland in the year 1980
A: 45.26962827644094

Q: what is the maximum number of male workers employed for the countries france, finland in the year 1980
A: 48.97010591234

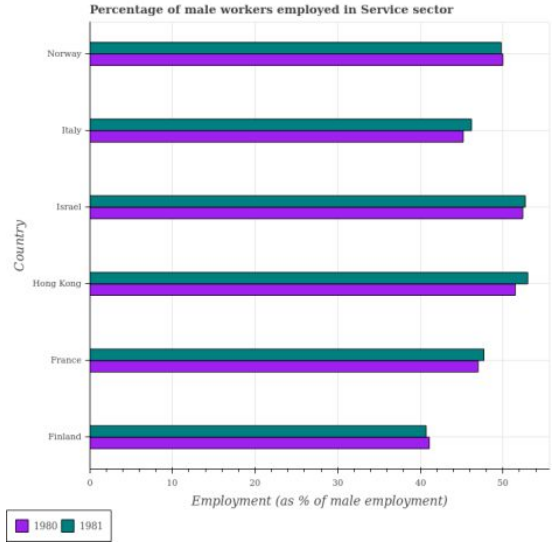
13. Project operation on column

Q: what are the names all the countries
A: Finland, France, Hong Kong, Israel, Italy, Norway

14. Range

Q: what is the range of % of male employment for the year 1980
A: RANGE = 12.622419158216474

TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

15. Quartiles (Q1 and Q3)

Q: find the quartiles for the year 1980
A: FIRST QUARTILE (Q1) = 43.97133559037385
SECOND QUARTILE (Q2) = 50.22946206732409
THIRD QUARTILE (Q3) = 53.90779749715967

16. IQR

Q: find the interquartile range for the year 1980
A: INTER-QUARTILE RANGE = 9.936461906785823

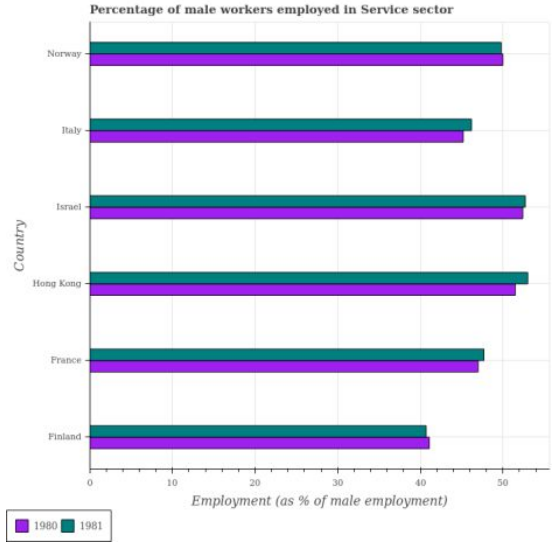
17. Structural Query

Q: what is the title of the graph ?
A: TITLE OF THE GRAPH = Percentage of male workers employed in Service se

Q: what is the label or title of the x-axis ?
A: X-LABEL = Employment (as % of male employment)

Q: what is the label or title of the y-axis ?
A: Y-LABEL = Country

TYPES OF QUESTIONS ADDRESSED



	Country	1980	1981
0	Israel	54.191570	54.512780
1	France	48.970106	48.401403
2	Finland	41.569151	41.217884
3	Norway	51.488818	51.018932
4	Italy	46.373521	46.332053
5	Hong Kong	53.624025	54.126604

18. YES / NO

Q: in the year 1981, hong kong has the highest employment percentage.
A: YES

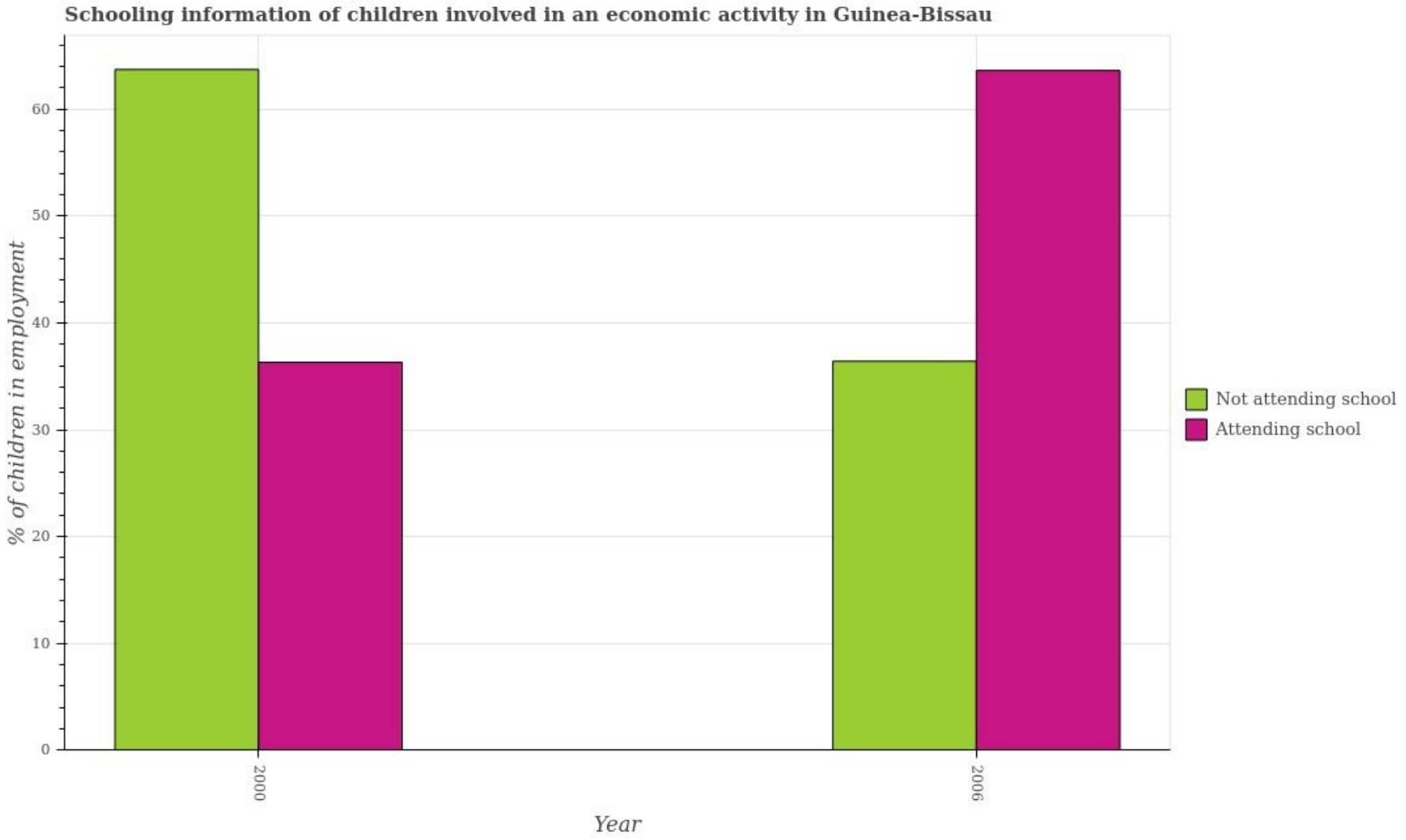
Q: for the year 1980, italy has a lower employment percentage than norway. A: YES

Q: the average employment rate of hong kong is greater than the average employment rate of finland.
A: YES

Q: is the percentage of male workers employed in service sector in 1980 in italy less than that in norway ?
A: YES

Project Demonstration - An Example

An Input to the system



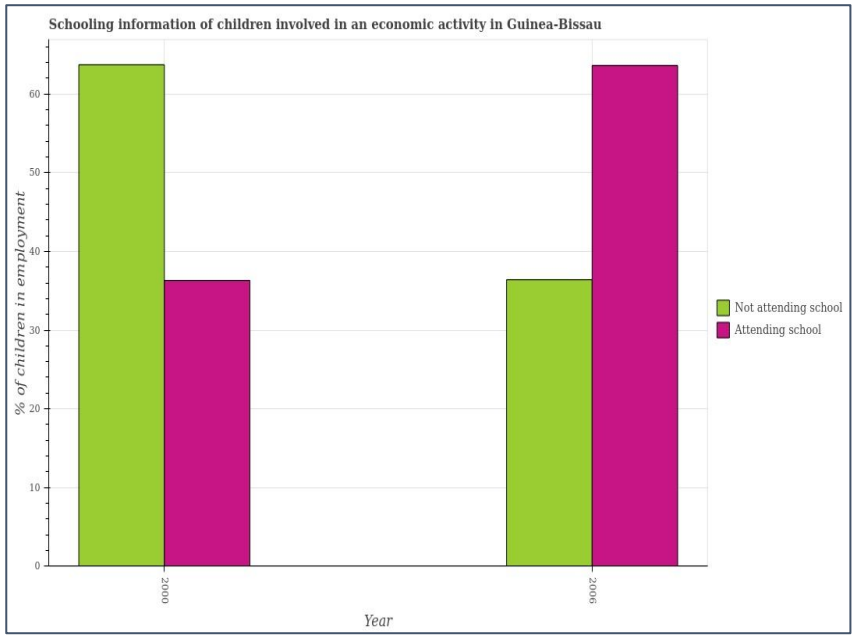
A Vertical Grouped Bar graph

Project Demonstration - An Example

Category	Score	left_x	top_y	right_x	bottom_y
yticklabel	0.9999904632568359	27.946645736694336	246.9253387451172	41.919342041015625	260.8976135253906
yticklabel	0.9999895095825195	27.873197555541992	328.9063720703125	41.86872100830078	343.05303955078125
bar	0.9999873638153076	640.291259765625	284.0574645996094	749.989990234375	580.9351806640625
bar	0.9999823570251465	201.92819213867188	281.8130798339844	312.2918701171875	578.4205322265625
legend_label	0.9999779462814331	932.5986328125	303.1197509765625	1071.5126953125	322.6226806640625
yticklabel	0.9999773502349854	27.868160247802734	409.8987121582031	41.989967346191406	423.9773254394531
bar	0.9999696016311646	750.8607177734375	62.46775817871094	861.982421875	581.0493621826172
xticklabel	0.9999669790267944	194.113037109375	590.9034423828125	208.31503295898438	618.8867797851562
yticklabel	0.9999643564224243	34.982704162597656	572.9815063476562	41.99324035644531	587.0731201171875
bar	0.9999592304229736	91.97103881835938	59.920265197753906	202.5250701904297	581.5347671508789
preview	0.9999498128890991	908.1177368164062	303.0318908691406	928.0110473632812	322.7835998535156
yticklabel	0.9999486207962036	28.012062072753906	164.76251220703125	42.00925827026367	178.90281677246094
yticklabel	0.9999397993087769	28.047203063964844	492.2265625	42.02458572387695	505.870849609375
xlabel	0.9999388456344604	457.864501953125	627.0880737304688	493.36260986328125	646.0440673828125
xticklabel	0.999916672706604	743.106201171875	590.95849609375	756.9241943359375	619.1182861328125
legend_label	0.9998894929885864	932.1422119140625	325.8930969238281	1043.439697265625	345.8262023925781
preview	0.9998648166656494	908.1642456054688	326.23406982421875	928.19921875	346.04678344726562
yticklabel	0.9998058676719666	28.096742630004883	84.25582885742188	42.00232696533203	97.83050537109375
ylabel	0.9998049139976501	6.735439777374268	190.0906524658203	26.02379274368286	425.5273742675781
title	0.9994369149208069	49.137367248535156	9.029414176940918	796.2882461547852	27.020319938659668

- Output of PED stage
- A json converted to txt
- Input for OCR stage

Project Demonstration - An Example

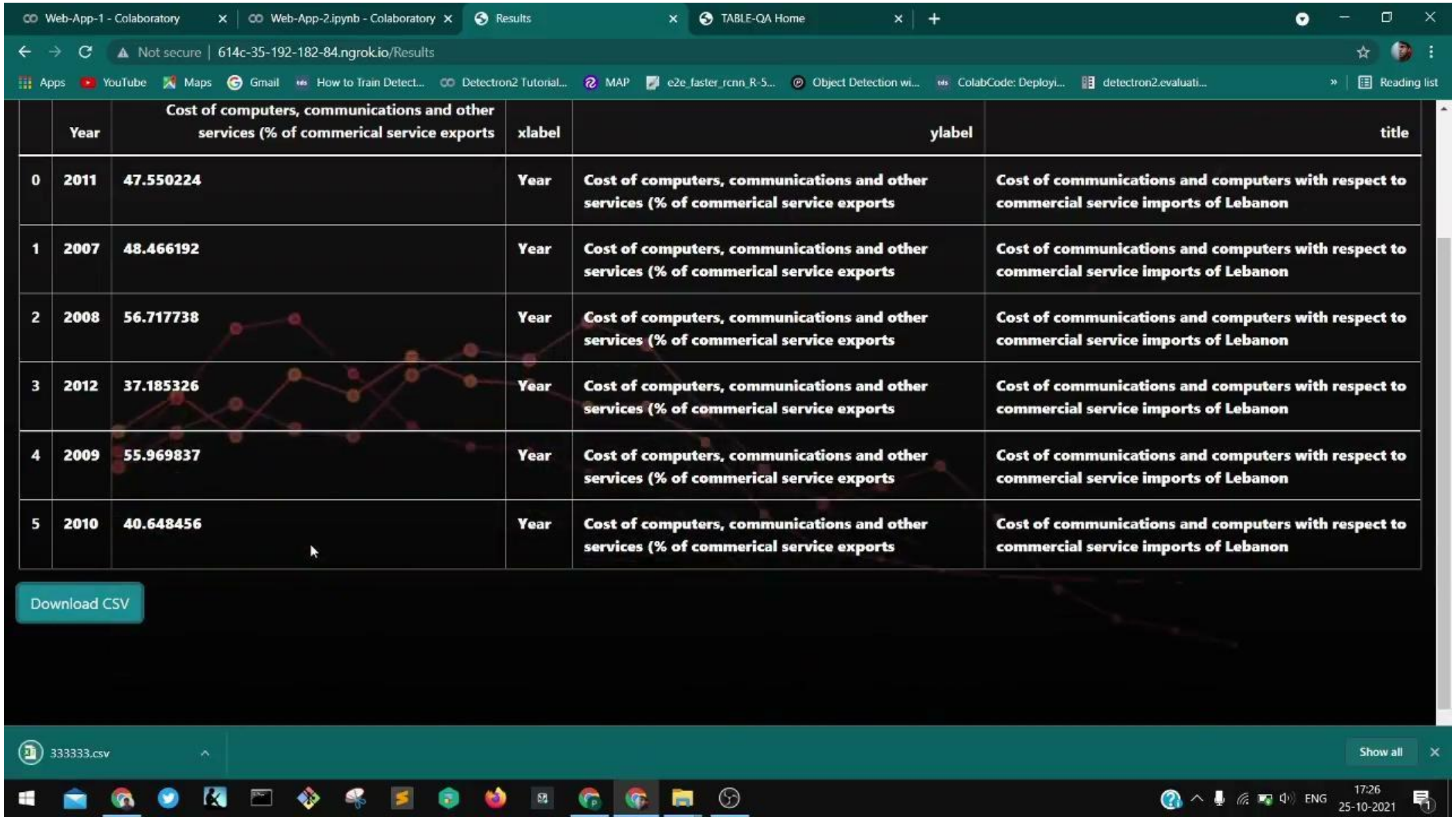


	Year	Attending school	Not attending school
0	2000	37.06615560158244	64.78018619662012
1	2006	64.48456969920525	37.0474010349917

CSV Equivalent Of The Input Plot

```
> which year has the highest % of children Not attending school
2006
> find the rat > find the median of % of children in Attending school category  the year 2000
RATION =  0.57 MEDIAN =  50.775362650393845
                                                                              00?
> which year has the lowest % of children Attending school
2000
```

Walkthrough



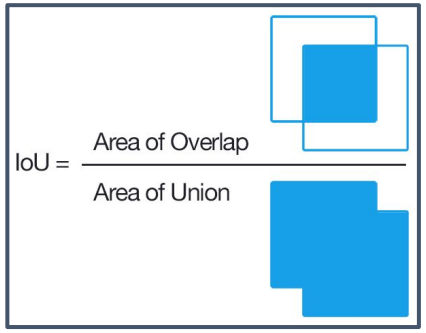
	Year	Cost of computers, communications and other services (% of commerical service exports	xlabel	ylabel	title
0	2011	47.550224	Year	Cost of computers, communications and other services (% of commerical service exports	Cost of communications and computers with respect to commercial service imports of Lebanon
1	2007	48.466192	Year	Cost of computers, communications and other services (% of commerical service exports	Cost of communications and computers with respect to commercial service imports of Lebanon
2	2008	56.717738	Year	Cost of computers, communications and other services (% of commerical service exports	Cost of communications and computers with respect to commercial service imports of Lebanon
3	2012	37.185326	Year	Cost of computers, communications and other services (% of commerical service exports	Cost of communications and computers with respect to commercial service imports of Lebanon
4	2009	55.969837	Year	Cost of computers, communications and other services (% of commerical service exports	Cost of communications and computers with respect to commercial service imports of Lebanon
5	2010	40.648456	Year	Cost of computers, communications and other services (% of commerical service exports	Cost of communications and computers with respect to commercial service imports of Lebanon

Test Plan and Strategy - PED Stage

PLOT ELEMENT DETECTION STAGE - EVALUATION USING DETECTRON-2

- Since Detectron-2 is an object detection tool, we make use of AP(Average Precision) as the evaluation metric.
- **AP** is the area under the PR Curve (Precision - Recall Curve) for an IoU threshold alpha.

IoU =



Threshold : alpha

We calculate the area of overlap and union between the predicted bbox and the actual expected bbox. We then calculate the AP for an IoU of 0.5.

True Positive (TP) — A valid detection. False Positive (FP) — An invalid detection. False Negative (FN) — Ground-truth missed by the model. For a perfect object detection model P = 1 and R = 1

Test Plan and Strategy - TABLE_QA STAGE

TABLE QUESTION ANSWERING - EVALUATION OF TaPas MODEL

- The predicted answer is evaluated against the expected answer of the PlotQa dataset.
- A textual answer is considered correct if it exactly matches the expected answer
- A numeric floating point answer is considered correct if it lies within 5% of the expected answer i.e.

expected answer +- (5% of the expected answer)

$$PerImageAccuracy = \frac{\sum CorrectlyAnsweredQuestions}{\sum QuestionsPosed}$$

$$OverallAccuracy = \frac{\sum PerImageAccuracy}{\sum ImagesinDataset}$$

Results - PED STAGE

- Below we present the figures we obtained on testing of the Detectron - 2 model
- AP [Average Precision] is the metric used for evaluation .
- Higher the AP , higher the area of overlap between actual and predicted boxes.

Number of Iterations	AP	AP50	AP75
100,000	79.621	91.956	90.397
130,000	86.584	92.819	92.076
150,000	87.014	92.825	92.086
170,000	87.053	92.823	92.083
200,000	87.179	92.823	92.088

Class	AP
Bar	88.819
Line	61.466
Dot-Line	77.439
X-Label	97.840
Y-Label	98.438
Title	90.056
X-tick Label	96.500
Y-tick Label	71.094
Legend Label	95.550
Preview	94.589

Results and Discussion - Table-QA Stage

- Accuracy as evaluation metric
- Textual answers = **Exact Match**
- Floating-point answers = **Within 5% of the expected answer**

Plot Type	Number of Images Tested	Total Number of Questions	Number of Correct Answers	Average Accuracy (in %)
Dot	2000	53970	25104	**46.965499**
Vertical	2000	47940	19898	**41.474200**
Horizontal	2000	49241	20128	**40.990114**
Line	2000	35353	14077	**36.669402**

Schedule

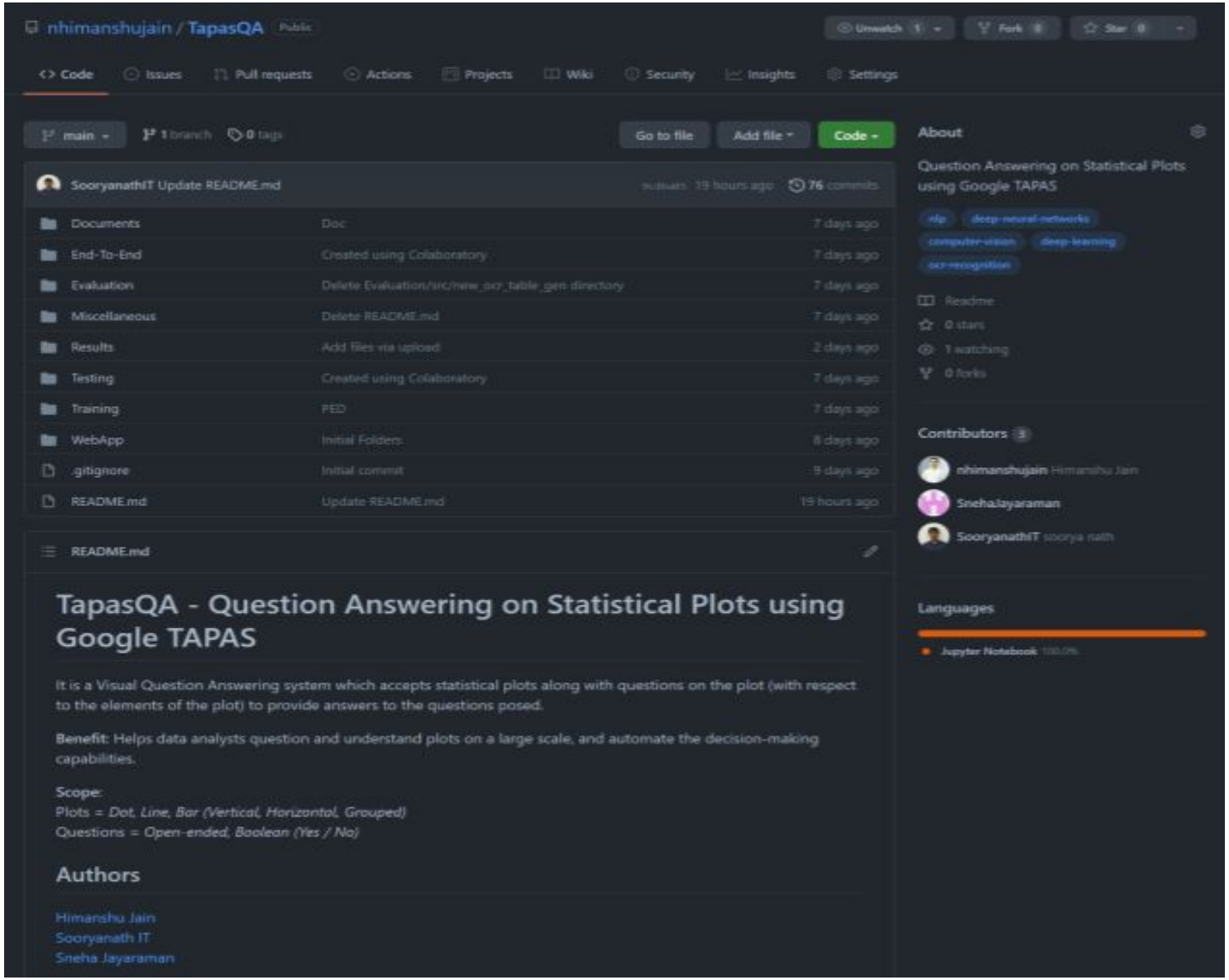
Task Tracker

Id	Level	Name	Progress	Start	Due	User
1	1	Dry Run Of Existing Model	100%	08/14/2021	08/19/2021	Sneha Jayaraman & Sooryanath I T
2	1	Identifying the Stages + Initiation of PED Stage - Closure	100%	08/19/2021	09/19/2021	Sooryanath I T
3	1	VED-OCR Stage Initiation - Completion	100%	08/28/2021	09/30/2021	Sneha Jayaraman
4	1	Capstone Review - 1	100%	08/29/2021	09/03/2021	Common
5	1	Table-QA Stage Initiation:Closure	100%	09/13/2021	10/15/2021	Himanshu Jain
6	1	Capstone Review - 2	100%	09/24/2021	09/26/2021	Common
7	1	Documentation Of Deliverables	100%	10/08/2021	11/03/2021	Common
8	1	Capstone Review - 3	100%	10/25/2021	10/29/2021	Common
9	1	ESA-Review	88%	10/29/2021	12/14/2021	Common
9.1	2	Automated - Evaluation	100%	10/29/2021	11/18/2021	Himanshu Jain
9.2	2	Capstone ESA Review	0%	12/11/2021	12/14/2021	Common

Documentation

- Project report finalized by Guide?
  - **Yes**
  - **Project_Report**
- IEEE (similar) Format of Paper ready for submission or current status?
  - **Paper is ready, post discussion with guide, it will be submitted to the conference.**
- Video (2-3 minutes) of your project? Please Play.
  - **PW22MHR02.mp4**
- All artifacts of your project uploaded in the CSE Project repository?
  - **We'll upload post this presentation .**

Documentation - Github Repository



Documentation - Project Poster

PES UNIVERSITY

VISUAL QUESTION ANSWERING ON STATISTICAL PLOTS

Team Members: Sneha Jayaraman, Sooryanath I T, Himanshu Jain
Guide: Dr. Mamatha H.R
Department of Computer Science and Engineering, PES University

Problem Statement

To build a **Visual Question Answering system** which accepts **statistical plots** along with **questions** on the plot (with respect to the elements of the plot) to provide answers to the questions posed.

Benefit: Helps data analysts question and understand plots on a large scale, **and automate the decision-making capabilities.**

Scope: **Plots = Dot, Line, Bar (Vertical, Horizontal, Grouped), Questions = Open-ended, Boolean (Support/ Refute)**

Background

Statistical plots consist of **Visual Elements** (Bars, lines, dots) and **Textual Elements** (x-label, y-label, title, x-tick, y-tick). We need to extract information out of visual elements and textual elements to create a semi-structured table that can then be used to answer to questions.

Dataset and Features

The dataset used is **PlotQA**. It consists of images with corresponding annotations (bounding boxes of elements), and question-answer pairs.

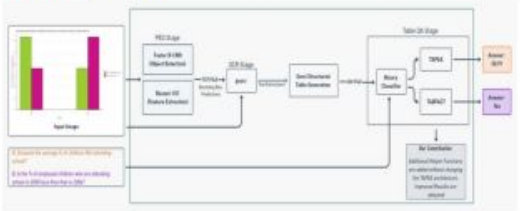
Requirements: Image of statistical plot and a question

Product Features: The product can answer **structural questions, data-retrieval, reasoning type and questions requiring a Yes/No answer** on any dot, line or bar plot.

Design Approach/ Methods

Approach: The design consists of 4 stages

1) Plot Element Detection Stage
2) Optical Character Recognition Stage
3) Semi-Structured Table Generation Stage
4) Table Question Answering Stage – Uses **TAPAS** model to answer open-ended questions, **TABFACT** to answer Yes/No questions



Results

Evaluation: Exact match for textual answers, 5% buffer for numeric answers.

Testing was done on a total of 8000 images and corresponding 1.86 Lakh question-answer pairs.

Plot Type	Number of Images Tested	Total Number of Questions	Number of Correct Answers	Average Accuracy (in %)
Dot	2000	53970	25104	**46.965499**
Vertical	2000	47940	19898	**41.474200**
Horizontal	2000	49241	20128	**40.990114**
Line	2000	35353	14077	**36.669402**

Summary of Project Outcome

Our model is able to answer questions posed on different types of plots significantly better using the TAPAS model with an **overall accuracy of 41.52%.**

It can answer questions requiring textual as well as numeric answers on different types of plots (hand-drawn and computer generated plots) including grouped plots.The product comes with an interactive user interface.

Conclusions and Future Work

This work is a step towards **machine reasoning capabilities**.

Future Work:

1) Extension to other types of plots (ex: pie etc)
2) Expansion to many different types of questions
3) Improving accuracy to match human accuracy.

References

Methani, Nitesh, et al. "Plotqa: Reasoning over scientific plots." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.

Kim, Dae Hyun, Enamul Hoque, and Maneesh Agrawala. "Answering questions about charts and generating visual explanations." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.

Sneha Jayaraman | Sooryanath I T | Himanshu Jain | Dr. Mamatha H.R

Lessons Learnt

- The learning curve throughout the process was steep, we had an opportunity to expose ourselves to various tools and practices.
- The team bonding added to accomplishing goals within deadlines in a hassle free manner.
- Venturing into a exciting problem statement helped us to explore various solutions and ultimately led us to developing on of our own.
- Regular meetings with our guide, helped us to bring the best output.
- The constructive criticism and feedback along with continual encouragement from the panel of professors helped us to correct ourselves and modify our style of working accordingly.

Conclusion and Future work

This work is a step towards **machine reasoning capabilities**.

Future Work:

1) Extension to other types of plots (ex: pie etc)
2) Expansion to many different types of questions
3) Improving accuracy to match human accuracy.

References

[1] R. Reddy, R. Ramesh, A. Deshpande, and M. M. Khapra, "Figurenet: A deep learning model for question-answering on scientific plots," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[2] N. Methani, P. Ganguly, M. M. Khapra, and P. Kumar, "Plotqa: Reasoning over scientific plots," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1527–1536.

[3] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[5] R. Smith, "An overview of the tesseract ocr engine," in *Ninth international conference on document analysis and recognition (ICDAR 2007)*, vol. 2. IEEE, 2007, pp. 629–633.

[6] J. Herzig, P. K. Nowak, T. Müller, F. Piccinno, and J. M. Eisenschlos, "Tapas: Weakly supervised table parsing via pre-training," *arXiv preprint arXiv:2004.02349*, 2020.

[7] P. Pasupat and P. Liang, "Compositional semantic parsing on semi-structured tables," *arXiv preprint arXiv:1508.00305*, 2015.

[8] J. C. Y. Z. H. W. S. L. X. Z. Wenhu Chen, Hongmin Wang and W. Y. Wang, "Tabfact : A large-scale dataset for table-based fact verification," in *International Conference on Learning Representations (ICLR)*, Addis Ababa, Ethiopia, April 2020.

[9] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

Thank You