



Document Layout Analysis

Problem Statement & Solution

Problem Statement:

Develop an accurate document layout analysis model to automatically identify and categorize textual and non-textual elements within document images. The model should handle diverse document types, languages, and layouts, outputting a structured representation of the layout.

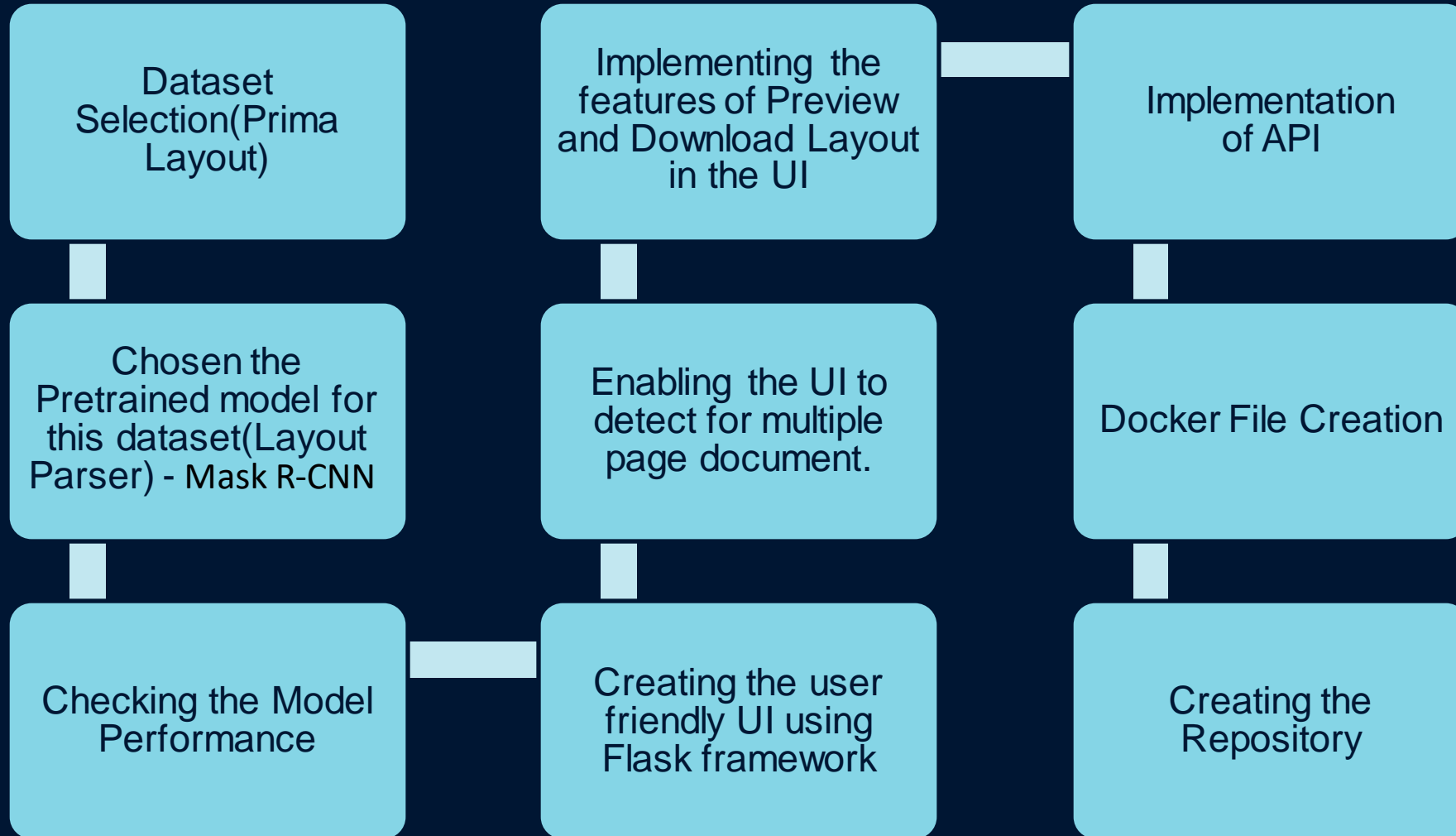
Solution:

The problem statement has been addressed through two solutions: a user-friendly Flask web application and a versatile Flask RESTful API. The web application enables seamless upload of PDFs or images, leveraging the Detectron2 layout analysis model to identify and visualize layout elements. This interactive interface allows users to understand the spatial arrangement of text regions, images, and tables in their documents. On the other hand, the RESTful API provides layout analysis functionality for PDFs and images, returning JSON responses with annotated bounding boxes. This API facilitates easy integration with various applications and services, ensuring scalability and adaptability. Both solutions empower users to gain valuable insights into their document layouts, fostering optimized OCR, document processing, and information extraction workflows.

Steps Involved:

- 1. Dataset Selection:** The Prima Dataset was chosen for its diverse content, which includes the scanned and native images/pdfs.
- 2. Model Selection:** Chosen a pretrained layout analysis model(layout parser trained using Detectron2), that can accurately detect and classify text regions, images, tables, and other elements.
- 3. Model Training and Evaluation:** We can train the model if we have the enough computational resources.
- 4. Visualization:** Visualize the detected layout elements by drawing bounding boxes around them on the document images.
- 5. Web Application Development:** Created a user-friendly Flask web application allowing users to upload PDFs or images for layout analysis.
- 6. API Implementation:** Set up a Flask RESTful API to provide layout analysis functionality for external applications and services.
- 7. Integration & Testing:** Integrate the model into the web application and API, ensuring smooth functionality and conducting thorough testing.
- 8. Deployment:** Deploy both the web application and RESTful API to suitable hosting environments for public access.
- 9. Documentation & Maintenance:** Create comprehensive documentation and establish a maintenance plan for long-term sustainability.

Use case – Process Flow



Results

Document Layout Analysis

Upload a PDF or an image to analyze the layout

Choose file No file chosen Analyze

S. No	File Name	No. of Pages	Preview	Date & Time	Download Layout
1	1.pdf	14	View	2023-09-28 16:19:49	Download
2	DocScanner 11 Jul 2023 17:54 (1).pdf	1	View	2023-09-29 03:30:14	Download
3	Test.jpeg	1	View	2023-09-29 03:30:33	Download

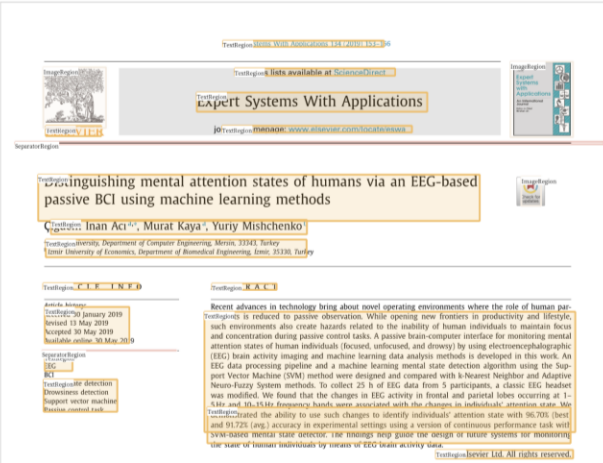
Clear Data

UI of Main Page

Layout Preview

File Name: 1.pdf

Page 1



UI of Layout Preview

Results - Cont.

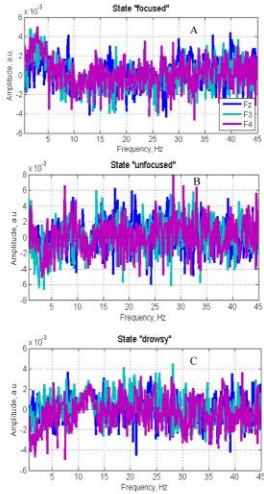


Fig. 5. An example of SVM weight vectors for focused (A), unfocused (B) and drowsy (C) mental states.

2.6. Performance evaluation

The performance of the SVM-based mental state detector is evaluated using random hold-out cross-validation. Cross-validation is a standard approach for evaluating the accuracy of classification and regression models in machine learning. In cross-validation, a subset of available data is withheld from training so that the machine learning model cannot see that data or have the knowledge of the information contained in the withheld dataset. After the training has finished, using whichever machine learning algorithm, an unbiased estimate of the performance of the final model was obtained by applying it to the withheld validation dataset and evaluating the accuracy of the model there. In random hold-out cross-validation, a certain percentage of randomly chosen data points was chosen and withheld prior to training and, subsequently, used for evaluation of the accuracy of the final detector.

Here, two evaluation paradigms (i.e. subject-specific and common-subject) were evaluated. In the subject-specific paradigm, the mental state detector was trained individually for each participant based on the data collected for that participant only. 80% of random time points from the EEG data from all experiments of that individual were used for training the SVM state-classifiers. The accuracy of the produced individual detector was then evaluated on the remaining 20% of the data of all trials that was not used in the training. Here, the training data of each participant generally consisted of 6000 data points randomly pulled together from different trials of the same subject. Respectively, 1500 data points were used for the evaluation of the performance of the final detector. In this approach, a different mental state detector was trained for each participant.

In the case of the common-subject paradigm, a single detector was jointly trained for all participants. 80% of the data from all joint EEG recordings of all participants was randomly selected for training. The performance of the "mixed" or "generic" detector was then evaluated against the dataset of mixed data, as well as against the data of each participant individually, which was not seen in the training. The cross-validation here was performed both jointly for all participants and individually for each participant to obtain the goodness of the generic mental state detector both overall, as well as on per-participant basis.

3. Results

All experiments in this study were performed with healthy volunteer participants chosen among students. All participants signed the informed consent form after receiving the instructions about the objectives and procedures of the experiments.

Fig. 6 shows a sample EEG data before the SVM-based classifier training part. The differences in the EEG signals associated with the different attention states can be seen in the spectrograms. Red dashed lines indicate the different attention state periods corresponding to the time-intervals of 0–10 min (focused), 10–20 min (unfocused), and 20–35 min (drowsy).

The preliminary examination of the collected data revealed that the changes in the individuals' attention states could be related to the changes in the EEG spectral power distribution, appearing as changes at certain frequencies in certain EEG channels. The varying composition of EEG signals with respect to the location of the electrodes can be seen in Fig. 7. In the figure, sections (A) and (B) show sample EEG spectrograms collected during the continuous performance task from the EEG electrodes Fz (A) and Pz (B) for participant S1. Section (C) indicates EEG spectrogram collected for participant S2 from the electrode Fz.

More specifically, focused and unfocused attention states appeared to be associated with enhanced or suppressed EEG activity at 1–10 Hz frequency bands in the frontal EEG channels F3, F4, and Fz. The drowsing state could be observed as continuous or intermittent power spouts at alpha-band frequency in the EEG signals at 10–15 Hz and the EEG channels C3, C4, Cz, and Pz. This examination suggests that focused, unfocused, and drowsing states of the participants could be distinguished in the EEG data based on the local spectrum of the signal in various EEG channels.

We then constructed an SVM-based algorithm for the detection of the above mental states from the EEG data, as described in Section 2.5. The examination of the spectral power features in such a detector supported the preliminary observations above. In particular, Table 2 lists some of the features of the mental state detector and its statistical properties, organized according to the intra class Correlation Coefficient (ICC) of such features with the target detector states. ICC is a statistical measure of relatedness of a continuous predictor variable with a discrete outcome variable.

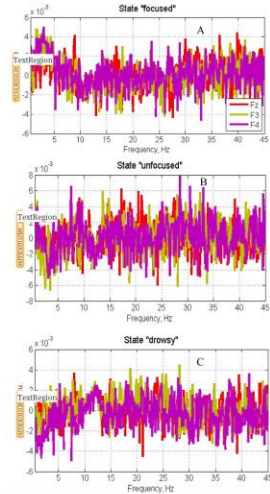


Fig. 6. Sample EEG data before the SVM-based classifier training part.

3.1. Performance evaluation

The performance of the SVM-based mental state detector is evaluated using random hold-out cross-validation. Cross-validation is a standard approach for evaluating the accuracy of classification and regression models in machine learning. In cross-validation, a subset of available data is withheld from training so that the machine learning model cannot see that data or have the knowledge of the information contained in the withheld dataset. After the training has finished, using whichever machine learning algorithm, an unbiased estimate of the performance of the final model was obtained by applying it to the withheld validation dataset and evaluating the accuracy of the model there. In random hold-out cross-validation, a certain percentage of randomly chosen data points was chosen and withheld prior to training and, subsequently, used for evaluation of the accuracy of the final detector.

Two evaluation paradigms (i.e. subject-specific and common-subject) were evaluated. In the subject-specific paradigm, the mental state detector was trained individually for each participant based on the data collected for that participant only. 80% of random time points from the EEG data from all experiments of that individual were used for training the SVM state-classifiers. The accuracy of the produced individual detector was then evaluated on the remaining 20% of the data of all trials that was not used in the training. Here, the training data of each participant generally consisted of 6000 data points randomly pulled together from different trials of the same subject. Respectively, 1500 data points were used for the evaluation of the performance of the final detector. In this approach, a different mental state detector was trained for each participant.

In the case of the common-subject paradigm, a single detector was jointly trained for all participants. 80% of the data from all joint EEG recordings of all participants was randomly selected for training. The performance of the "mixed" or "generic" detector was then evaluated against the dataset of mixed data, as well as against the data of each participant individually, which was not seen in the training. The cross-validation here was performed both jointly for all participants and individually for each participant to obtain the goodness of the generic mental state detector both overall, as well as on per-participant basis.

3.2. Results

All experiments in this study were performed with healthy volunteer participants chosen among students. All participants signed the informed consent form after receiving the instructions about the objectives and procedures of the experiments.

Fig. 7 shows a sample EEG data before the SVM-based classifier training part. The differences in the EEG signals associated with the different attention states can be seen in the spectrograms. Red dashed lines indicate the different attention state periods corresponding to the time-intervals of 0–10 min (focused), 10–20 min (unfocused), and 20–35 min (drowsy).

The preliminary examination of the collected data revealed that the changes in the individuals' attention states could be related to the changes in the EEG spectral power distribution, appearing as changes at certain frequencies in certain EEG channels. The varying composition of EEG signals with respect to the location of the electrodes can be seen in Fig. 7. In the figure, sections (A) and (B) show sample EEG spectrograms collected during the continuous performance task from the EEG electrodes Fz (A) and Pz (B) for participant S1. Section (C) indicates EEG spectrogram collected for participant S2 from the electrode Fz.

More specifically, focused and unfocused attention states appeared to be associated with enhanced or suppressed EEG activity at 1–10 Hz frequency bands in the frontal EEG channels F3, F4, and Fz. The drowsing state could be observed as continuous or intermittent power spouts at alpha-band frequency in the EEG signals at 10–15 Hz and the EEG channels C3, C4, Cz, and Pz. This examination suggests that focused, unfocused, and drowsing states of the participants could be distinguished in the EEG data based on the local spectrum of the signal in various EEG channels.

We then constructed an SVM-based algorithm for the detection of the above mental states from the EEG data, as described in Section 2.5. The examination of the spectral power features in such a detector supported the preliminary observations above. In particular, Table 2 lists some of the features of the mental state detector and its statistical properties, organized according to the intra class Correlation Coefficient (ICC) of such features with the target detector states. ICC is a statistical measure of relatedness of a continuous predictor variable with a discrete outcome variable.

WRITING

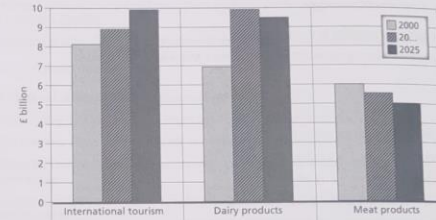
WRITING TASK 1

You should spend about 20 minutes on this task.

The chart below gives information about Southland's main exports in 2000, *20..., and future projections for 2025. Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

Southland's main exports in 2000 and *20..., and future projections for 2025



(*20... : for test purposes, use and refer to the current year)

WRITING TASK 2

You should spend about 40 minutes on this task.

In some countries an increasing number of people are suffering from health problems as a result of eating too much fast food. It is therefore necessary for governments to impose a higher tax on this kind of food. To what extent do you agree or disagree with this opinion?

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

WRITING

WRITING TASK 1

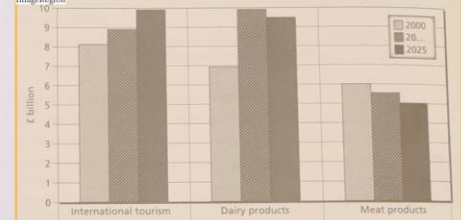
You should spend about 20 minutes on this task.

The chart below gives information about Southland's main exports in 2000, *20..., and future projections for 2025.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

Southland's main exports in 2000 and *20..., and future projections for 2025



(*20... : for test purposes, use and refer to the current year)

WRITING TASK 2

You should spend about 40 minutes on this task.

In some countries an increasing number of people are suffering from health problems as a result of eating too much fast food. It is therefore necessary for governments to impose a higher tax on this kind of food.

To what extent do you agree or disagree with this opinion?

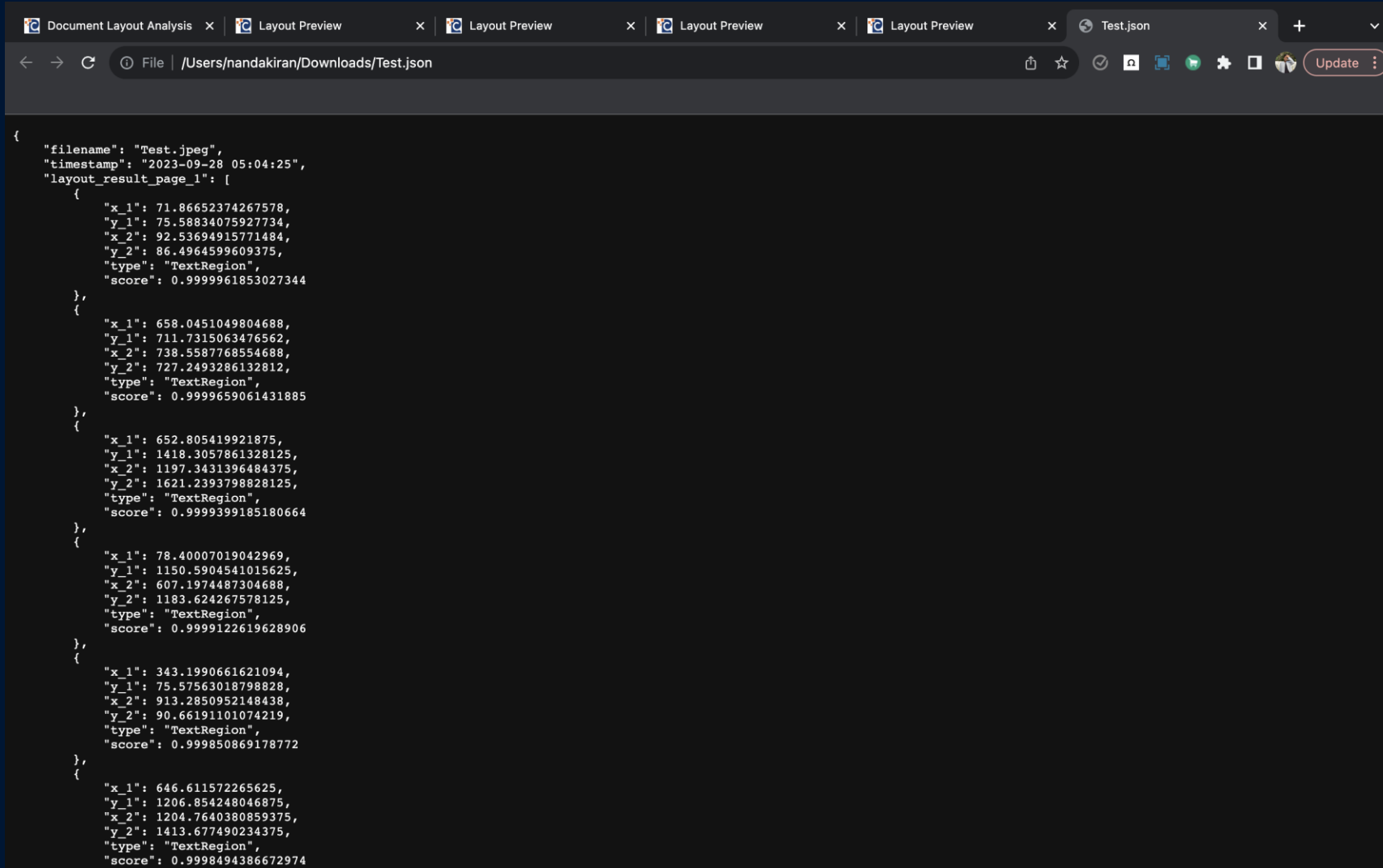
Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Native Image and Its Output

Scanned Image and Its output

Results- Cont.



```
{
  "filename": "Test.jpeg",
  "timestamp": "2023-09-28 05:04:25",
  "layout_result_page_1": [
    {
      "x_1": 71.86652374267578,
      "y_1": 75.58834075927734,
      "x_2": 92.53694915771484,
      "y_2": 86.4964599609375,
      "type": "TextRegion",
      "score": 0.9999961853027344
    },
    {
      "x_1": 658.0451049804688,
      "y_1": 711.7315063476562,
      "x_2": 738.5587768554688,
      "y_2": 727.2493286132812,
      "type": "TextRegion",
      "score": 0.9999659061431885
    },
    {
      "x_1": 652.805419921875,
      "y_1": 1418.3057861328125,
      "x_2": 1197.3431396484375,
      "y_2": 1621.2393798828125,
      "type": "TextRegion",
      "score": 0.9999399185180664
    },
    {
      "x_1": 78.40007019042969,
      "y_1": 1150.5904541015625,
      "x_2": 607.1974487304688,
      "y_2": 1183.624267578125,
      "type": "TextRegion",
      "score": 0.9999122619628906
    },
    {
      "x_1": 343.1990661621094,
      "y_1": 75.57563018798828,
      "x_2": 913.2850952148438,
      "y_2": 90.66191101074219,
      "type": "TextRegion",
      "score": 0.999850869178772
    },
    {
      "x_1": 646.611572265625,
      "y_1": 1206.854248046875,
      "x_2": 1204.7640380859375,
      "y_2": 1413.677490234375,
      "type": "TextRegion",
      "score": 0.9998494386672974
    }
  ]
}
```

JSON Output Format

Benefits

- **Improved Document Understanding:** Layout analysis provides a comprehensive view of document structure, enhancing comprehension of content arrangement and relationships.
- **Efficient Information Extraction:** Accurate layout analysis streamlines data extraction, enabling faster and more precise retrieval of text, tables, and key information.
- **Optimized OCR Performance:** Precise text region detection improves OCR accuracy, reducing errors and enhancing the overall performance of optical character recognition.
- **User-Friendly Interactions:** Interactive web applications and APIs allow users to easily analyze document layouts, offering a seamless and intuitive user experience.
- **Versatility and Integration:** Utilizing diverse datasets, like Prima, ensures adaptability to different document types, while the RESTful API enables effortless integration with other applications and services.

Future Work

- Training the existing pre trained layout parser model with our custom dataset so to integrate smoothly in our pipeline.
- Training the model in our local Environment to have more access towards the code.

THANKS!

NandaKiran Velaga
nandakiran.velaga@criticalriver.com

