

Problem Statement: Marks Extractor

Business Challenge Description: How can we leverage AI to accurately extract information from Indian mark sheets.

A major problem that we face today is the inability to leverage data in scanned documents and images. The challenge is not just to extract data from scanned documents but also to extract it accurately. This becomes even more challenging when the documents have different formats and information to be extracted from them.

Expected Business Solution

The expected solution should be a prototype model that can be executed on at least 2-3 different types of mark sheets from various boards to extract information and should be easily extensible to support multiple formats in the future.

Tools & Technology

You are free to use any available tools for research and implementation

Resources/References

- <https://cloud.google.com/document-ai/docs>
- <https://cloud.google.com/vision/docs/ocr>
- <https://towardsdatascience.com/extracting-text-from-scanned-pdf-using-pytesseract-open-cv-cd670ee38052>