APPROACH EXPLANATION

For this project, I developed a Social Media Content Analyzer that extracts text from PDF and image files, allowing users to analyze their content. The application uses Tesseract.js for OCR to extract text from images (JPG, PNG, JPEG) and pdf-parse for text extraction from PDF files.

Frontend

The frontend is built with React and Vite for a fast, modern user experience. It offers users the option to upload files through drag-and-drop or file selection using React-Dropzone library. Once the file is uploaded, it is sent to the backend using Axios for processing.

Backend

The backend, created with Node.js and Express, uses Multer to handle file uploads. For PDFs, it processes the content using pdf-parse, while for images, Tesseract.js converts image content into text. The extracted text is then sent back to the frontend and displayed to the user.

Additionally, users can clear the uploaded file and reset the text output. This application focuses on usability, efficiency, and providing meaningful insights by allowing users to quickly extract and analyze text content from various file formats.

Deployment

Backend Deployment

- The backend is deployed on the Render as Web Service.
  https://smca-backend.onrender.com

Frontend Deployment

- The frontend was deployed on Render as a Static Site.

  https://smca-frontend.onrender.com