

Kanish Mahato

21103070

Report: Bank Statement PDF to Excel Converter

1. Introduction

The Bank Statement PDF to Excel Converter is a Streamlit-based application that allows users to upload bank statement PDFs and convert them into an Excel file. This tool extracts text-based data from the PDF and presents it in a structured format for financial analysis and record-keeping.

2. Objectives

- Automate the conversion of bank statements from PDF to Excel.
- Provide an easy-to-use interface for users.
- Enable structured financial data extraction.
- Support seamless downloading of converted files.

3. Technologies Used

- Python: Programming language used for implementation.
- Streamlit: Web framework for creating the user interface.
- pdfplumber: Library used for extracting text from PDFs.
- Pandas: Data manipulation and structuring.
- XlsxWriter: Converts extracted data into an Excel file.
- HTML/CSS: Used for UI enhancements.

4. Features

a) File Upload

- Users can upload a text-based PDF file.
- The application verifies the file format before processing.

b) PDF Data Extraction

- Uses pdfplumber to read the text from each page of the uploaded PDF.
- Extracts transaction details and other financial data.

c) Data Display

- Extracted data is displayed in a structured dataframe format using Pandas.
- Users can review the extracted information before downloading.

d) Excel Conversion

- The extracted data is converted to an Excel file.
- The Excel sheet is formatted for easy readability.
- A download button is provided for user convenience.

e) UI Enhancements

- Styled download button with color change on click.
- Loading spinner while extracting data to enhance user experience.

5. Installation Guide

a) Prerequisites

- **Install Python (3.7+ recommended).**

b) Install Dependencies

Run the following command in the terminal:

```
pip install streamlit pdfplumber pandas xlswriter
```

c) Running the Application

To start the application, execute:

```
streamlit run app.py
```

6. Input and Output Examples

a) Sample Input PDF

(Add screenshots or descriptions of a sample bank statement PDF file.)

b) Extracted Data Output

(Add a screenshot of the extracted data displayed in the Streamlit app.)

c) Final Excel Output

(Add a screenshot of the downloaded Excel file and its structure.)

7. Limitations

- The tool only works with text-based PDFs (does not support image-based scanned documents).
- Formatting of extracted data may vary based on the PDF structure.
- No OCR support for images within PDFs.

8. Future Improvements

- Implement OCR support for scanned PDFs.
- Add AI-based transaction categorization.
- Enable multi-page PDF handling with advanced formatting.
- Cloud storage integration for storing and accessing processed files.

9. Conclusion

The Bank Statement PDF to Excel Converter simplifies financial data extraction and enhances productivity by automating a previously manual process. With further improvements, it can become a powerful financial analysis tool.

GitHub Repository: <https://github.com/tron2003/minal>