PIP104 PROFESSIONAL PRACTICE-II

VIVA-VOCE

PROJECT TITLE: ANALYTICS OF SCANNED PRESCRIPTIONS AND NOTES.

Batch Number: 8

Roll Number	Student Name
20201CCS0058	KAVYA UNNIKRISHNAN
20201CCS0145	K P LAKSHMI SHARANYA
20201CCS0001	SAI HARSHITHA
20201CCS0018	KETHA RISHITHA

Under the Supervision of,

Dr. Nihar Ranjan Nayak
Professor / Associate Professor / Assistant Professor
School of Computer Science Engineering & Information Science
Presidency University

PRESIDENCY UNIVERSITY
Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

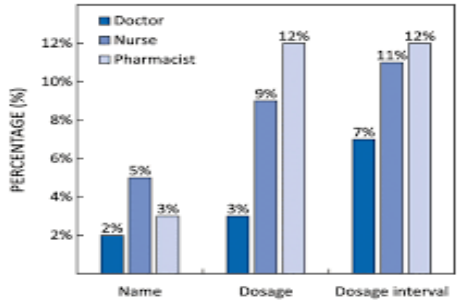
Introduction

- In today's digital world, healthcare is one of the core areas in the medical domain. A healthcare system is required to analyze a large amount of patient data, which helps to derive insights and predictions of disease. In the current day's healthcare system, the adaptation of digitization of medical data, such as data related to patients' diagnosis, generates a big volume of data in a short period of time. To analyze this huge volume of medical data, we need techniques that have the power of statistical analysis to predict or extract hidden important information out of the data. Therefore, the need for data analytics in healthcare imaging has increased more rapidly. In this chapter, an analytical study on big data analytics in healthcare is carried out. This study includes all possible terminology related to the idea of data extraction, healthcare data, and architectural context for data analytics, with different tools and platforms discussed in detail. A review of some recent noteworthy contributions in the field is also presented.

Literature Review

- Every visit to the doctor's office involves the generation of one or multiple prescriptions. For those who frequent the doctor's office often, the prospect of keeping track of yet another prescription is dreadful as it needs to be maintained for a long duration to avoid any side effects in case new medicines that have been prescribed generate an unforeseen reaction. From the point of view of the doctor or the hospital at large, they also need to maintain comprehensive records of the hundreds if not thousands of patients that visit them, this makes it exceptionally tedious to maintain all these records

Figure 1: Percentage of errors per prescription.



Literature Review

- Since time immemorial, the relationship between doctor and patient has included a prescription as an intermediary for the benefit of the patient. It is through this prescription that doctors communicate the remedy of the patient's ailment to not just the patient but also the pharmacist who is going to provide those medicines and to other doctors that the patient may ever visit. The problem arises when someone who needs to visit the doctor frequently has to keep track of their medicines and allergies. The higher the number of visits, the greater the volume of the paper trail that needs to be maintained. It slowly becomes a very cumbersome task to store all these documents as they become the backbone of the medical history and record of the patient. Even in the modernized 21st century most establishments still rely on traditional methods to generate these prescriptions

Research Gaps Identified

- Retaining the original formatting, such as tables or columns, is challenging in the extraction process
- PDFs can contain complex layouts, images, or non-standard text encoding, sometimes it resulted as inaccurate text extraction.
- Some libraries might not handle certain types of PDF layouts or versions effectively.
- For large PDFs or high volumes of documents, these methods might be resource-intensive or slow.
- Extracted data might require additional cleaning or post-processing due to extraction errors.
- Password-protected or encrypted PDFs might pose limitations in extracting content.

Proposed Methodology

- Data Sources and Collection
- Assess the PDF Structure
- Manual Extraction
- Automated Extraction
- Data Cleaning and Formatting
- Text Mining and Natural Language Processing
- Predictive Analytics and Decision Support
- Medication Adherence and Management
- Scrape PDF Data in Structured Form/Unstructured form
- Import PDF data as a DataFrame
- Create a Row Identifier using python libraries

Proposed Methodology

- Reshape the data (convert data from long-form to wide form)
- Join the data in the left section with the data in the right section
- Pattern Recognition and Trend Analysis
- Quality Assessment:
- Integration with Electronic Health Records (EHR)
- Privacy and Security Measures
- User Interface and Reporting

Objectives

- Automating the process of scraping data from PDF files and converting unstructured data into panel data can be a complex but valuable task. The objectives for this project could include:
- **Data Extraction**: Develop a system to extract data from PDF files, identifying key information such as text, tables, and images.
- **Text Parsing**: Convert extracted text into structured data by parsing it for relevant content, such as dates, numbers, and keywords.
- **Data Cleaning**: Implement data cleaning procedures to handle inconsistencies, missing values, and errors in the extracted data.
- **Data Integration**: Combine data from multiple PDF files into a coherent panel dataset, ensuring data consistency and accuracy.

Objectives

- **Panel Data Format**: Transform the structured data into panel data format, where data is organized in rows and columns, typically with a time dimension and cross-sectional units.
- **Data Storage**: Establish a database or file storage system to store the panel data securely.
- **Automation**: Create an automated workflow or script to regularly scrape and update the panel data from new PDF files as they become available.
- **Documentation:** Document the entire process, including data sources, extraction methods, data transformation, and analysis procedures for future reference and transparency.
- **Security and Compliance:** Ensure data security and compliance with relevant regulations, especially if the data involves sensitive or personal information.

Objectives

- **Error Handling:** Implement error handling and logging to identify and address issues in the automated process.
- **User-Friendly Interface:** If applicable, create a user-friendly interface for users to interact with the panel data and access insights.
- **Maintenance and Updates:** Plan for regular maintenance and updates to adapt to changes in data sources, file formats, or extraction techniques.
- These objectives will help the development of a robust and efficient system for automating the extraction and transformation of data from PDF files into panel data.

System Design & Implementation

System Design & Implementation

Timeline of Project

Week 1-2 - Research and data collection

Week 3-8 - Planning and Requirement Analysis

Week 9-10 - Testing and Evaluation

Week 10-12 - Data Validation and Verification

Week 12+ - Documentation and Reporting

Outcomes / Results Obtained

- **Text Extraction:**

  Extracting plain text from PDF documents.

- **Table Extraction:**

  Extracting tabular data from PDFs and organizing it into structured formats like CSV or Excel.

- **Key-Value Pair Extraction:**

  Extracting data in a key-value format from PDFs, where specific patterns or labels are associated with corresponding values.

- **Data Visualization:**

  Visualizing the extracted data to provide insights or facilitate understanding

Outcomes / Results Obtained

- **Information Summarization:**

  Summarizing extracted text to provide a concise representation of the key information within the PDF.

- **Data Cleaning and Preprocessing:**

  Cleaning and preprocessing the extracted data to remove noise, handle missing values, or standardize formats.

- **Data Export:**

  Exporting the extracted and processed data to various formats such as CSV, Excel, JSON, or a database.

Outcomes / Results Obtained

- **Error Handling and Logging:**

  Implementing robust error handling and logging mechanisms to track any issues during the extraction process

- **Scalability:**

  Designing the project in a way that allows for scalability, enabling efficient handling of a large volume of PDFs.

- **Named Entity Recognition (NER):**

  Identifying and extracting specific entities such as names, organizations, dates, etc. from the PDF text.

Conclusion

- Patient happiness and the quality of health care as a whole can be significantly impacted by the length of time they have to wait while receiving treatment.
- In conclusion, the analytics of scanned prescriptions and notes holds immense potential for transforming healthcare practices and improving patient outcomes. By leveraging advanced technologies such as optical character recognition (OCR) and natural language processing (NLP), healthcare providers can extract valuable insights from these documents.
- The systematic analysis of scanned prescriptions allows for the identification of patterns in medication prescriptions, dosage adjustments, and adherence trends. This information can contribute to more informed decision-making by healthcare professionals, leading to personalized and optimized treatment plans for patients.

Conclusion

- In this fast paced and modernized world, we need to keep up with the latest technology and implement that in the simplest of situations to replace traditional redundant methods, the method in question being converting paper based prescriptions to text based digital format documents. It is something that could revolutionize the way the medical field works and ensure smooth interaction among all parties involved in the process, be it the interaction between doctor and patient or patient and pharmacist, it assists in the smooth functioning of the process thus eliminating the need for maintaining physical records.

References

- Esraa Hassan, Habiba Tarek, Mai Hazem, Shaza Bahnacy, Lobna Shaheen and Walaa H. Elashmwai, “Medical Prescription Recognition using Machine Learning
- Zobeir Raisi, Mohamed A. Naiel, Steven Wardell, Paul Fieguth and John Zelek " Handwritten Recognition using Transformer for Medical Prescription Application." IEEE transactions on transformers, no. 1 (2019).
- Padmanabhan, Sharan, and Nuvvuru Divya. "Doctors Handwritten Prescription Recognition System In Multi Language Using Deep Learning." arXiv preprint arXiv:2210.11666 (2022).
- Sandhya, P., and K. P. Rama Prabha. "Comparison Of Various Machine Learning Algorithms For Recognizing Text On The Medical Prescriptions." Journal of Pharmaceutical Negative Results (2022): 2083-2091.

Publication Details

Achievements (if any)

Thank You