

Summary Report of Work Completed on Day 1

1. **Development of a Program to Save and Retrieve Data from PDFs for Further Processing (In Progress)**
 - Initiated the development of a Java program utilizing the Apache PDFBox library to extract text data from PDF documents.
 - Converted PDF content into structured text files for easier handling and future data manipulation.
2. **Study of DMC Documentation to Understand Key Variables and Data Requirements**
 - Reviewed DMC documentation to identify critical variables, data structures, and relationships between data elements.
 - Established a foundation for effective data handling and processing protocols.
3. **Review of Fundamental Concepts in Data Science, Including Data Mining and Machine Learning**
 - Focused on core data mining methods, including association rule mining, clustering, and classification.
 - Explored machine learning concepts, particularly supervised and unsupervised learning techniques, and reviewed essential terminology.
4. **Research into the Foundations of Natural Language Processing (NLP)**
 - Investigated fundamental NLP concepts such as tokenization, stemming, and lemmatization.
 - Examined NLP techniques and tools relevant for processing and analyzing textual data from PDFs.