

Software Engineering Internship at DataSoft Systems Bangladesh Limited

by

MD. Shadman Murshed
21301242

An internship report submitted to the Department of Computer Science and
Engineering in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2026

© 2026. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The internship report submitted is my/our own original work while completing degree at Brac University.
2. The report does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The report does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

MD. Shadman Murshed

21301242

Approval

The thesis/project titled “Software Engineering Internship at DataSoft Systems Bangladesh Limited ” submitted by

1. MD. Shadman Murshed (21301242)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in January 31, 2025.

Examining Committee:

Supervisor:
(Member)

Mr. Rafeed Rahman

Senior lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD

Professor
Department of Computer Science And Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

This report documents a comprehensive software engineering internship at Data-Soft Systems Bangladesh Limited. This internship involved the development of two primary projects, one is Training PRO and other is DocuCraft Pro. Training PRO is an asynchronous Learning Management System (LMS) that uses Redis for real time progress tracking across various media formats. DocuCraft Pro is an advanced document processing tool which provides multi-method architecture like No OCR, Tesseract based OCR and GenAI based extraction to handle the complex structure of Bangla documents, specifically converting Bijoy encoding into standard Unicode. Beside the projects, the internship experience included working with the Business Analyst team, gaining Software Quality Assurance (SQA) knowledge and exploring Android development with Java. Technical exposure was also included with DevOps, Docker, CI/CD pipelines and Nginx for load balancing. Throughout the internship, technical challenges for example facing API limits, handling complex PowerPoint structures are overcome through resourceful engineering. The internship helps the transition from theoretical academic knowledge to professional industry oriented real world full stack development, AI and API integration and collaborative software engineering within an Agile environment.

Keywords: Learning Management System (LMS), Redis, OCR, GenAI, Flask, Android Development, DevOps, Docker, SQA, NILA Chatbot.

Acknowledgment

I would like to begin by thanking Almighty Allah for providing me with the strength and patience to complete this internship successfully. I want to express my deepest gratitude to my supervisor Mr. Rafeed Rahman, Senior Lecturer at the Department of Computer Science and Engineering, Brac University. His mentorship, constant encouragement and valuable technical insights were instrumental in helping me navigate the various challenges of developing Training PRO and DocuCraft Pro. I also want to thank the entire team at DataSoft Systems Bangladesh Limited for giving me the opportunity to work in a professional environment and gain such diverse real world experience. Finally, I am forever indebted to my parents for their continuous support, love, and sacrifice throughout my academic journey. Their unwavering belief in my potential has been my greatest source of motivation and strength during every step of my life.

Table of Contents

Declaration	i
Approval	ii
Ethics Statement	iii
Abstract	iii
Dedication	iv
Acknowledgment	iv
Table of Contents	v
List of Figures	vii
Nomenclature	vii
1 Introduction	1
1.1 About Internship	1
1.2 About the Report	1
1.3 Problem Statement	1
1.4 Objectives	2
1.5 Methodology	2
1.6 Literature Review	2
2 Company Profile	3
2.1 Company Overview	3
2.2 Vision and Mission	3
2.2.1 Vision	3
2.2.2 Mission	3
2.3 Operations and Strategic Partnerships	4
2.4 Impactful Products and National Contributions	4
2.4.1 Rapid Pass (Metro Rail Integration)	4
2.4.2 Financial and Microfinance Solutions	4
2.4.3 Infrastructure and Port Automation	4
2.5 Software Development Process and Technology Stack	4
2.6 Company Culture and Values	5

3	Training and Development	6
3.1	Training and Hands-on Experience	6
3.2	Skill Development at DataSoft	6
3.3	Team Collaboration and bKash AML Team	6
3.4	Mentorship and Professional Guidance	7
3.5	Personal Growth and Professional Skills	7
4	Proposed Methodology	8
4.1	Design Process or Methodology Overview	8
4.2	Project 1: Training PRO - A LMS Platform	9
4.2.1	Detailed Implementation	9
4.2.2	Tools, Languages, and Frameworks Used	10
4.2.3	Purpose of the Project	10
4.2.4	Project Screenshots	11
4.3	Project 2: DocuCraft Pro: PDF Text Extraction System	13
4.3.1	Detailed Implementation	13
4.3.2	Project Screenshots	14
5	Challenges and Learning Outcomes	16
5.1	Technical Challenges	16
5.1.1	Challenges in Training PRO	16
5.1.2	Challenges in DocuCraft Pro	16
5.2	Technical Learning and Professional Growth	17
5.2.1	Full-Stack Development Mastery	17
5.2.2	AI and Document Processing Expertise	17
5.2.3	Experience with AI Chatbots and SQA	17
5.2.4	Android Development Exploration	17
5.2.5	DevOps and Infrastructure	17
5.3	Summary of Learning Outcomes	18
6	Conclusion	19
6.1	Summary of Findings	19
6.2	Professional Contributions	19
6.2.1	Collaboration with Business Analysts	19
6.2.2	Technical Presentations	19
6.2.3	Team Integration	20
6.3	Recommendations for Future Work	20
6.3.1	Future Plans for Training PRO	20
6.3.2	Future Plans for DocuCraft Pro	20
6.4	Concluding Remarks	20
	Bibliography	21

List of Figures

4.1	The Instructor main dashboard to manage courses.	11
4.2	The course management interface for an individual course.	11
4.3	The Student Learning Dashboard interface in Training PRO.	12
4.4	Example of Docx displaying the student’s scroll-based progress (32.0%)	12
4.5	The main interface for DocuCraft PRO.	14
4.6	A side by side comparison of the conversion by the Gemini AI model.	15

Chapter 1

Introduction

1.1 About Internship

In our university curriculum, students are given the choice between a project, a thesis, an internship. I choose to pursue an internship because I wanted to gain hands on experience in a professional setting rather than just focusing on theoretical research. My goal was to explore the software industry from the inside and become familiar with cutting edge technologies that are actually used in production level. Furthermore, my future plan is to become a software engineer and stay in my country. I believe that by working here I can do something impactful for the people of Bangladesh. Doing an internship at a prestigious company like DataSoft Systems Bangladesh Limited provided the perfect foundation for this career path.

1.2 About the Report

This report will give a detailed documentation of my experience as a Software Engineering Intern. It shows the technical and professional growth that I achieved while working on two major projects: **Training PRO** and **DocuCraft Pro**. The report will cover everything from the initial requirements analysis then system design to the final implementation also the challenges I faced during development. Moreover, it reflects the learning outcomes and how this experience helps to prepare myself for a professional role in the tech industry.

1.3 Problem Statement

During my internship, I worked with two primary Projects:

- **Fragmented Learning Systems:** Traditional LMS platforms often lack an end to end, lightweight environment for managing various content formats. Instructors face difficulties to track student progress accurately.
- **Bangla Document Digitization:** Existing PDF extraction tools most of the time fails to handle Bangla Bijoy formats or complex scanned layouts. It creates a significant barrier for organizations in Bangladesh.

1.4 Objectives

The core objectives of my internship were as follows:

- To build **Training PRO**, an asynchronous LMS with real time progress tracking feature for different types of media formats.
- To develop **DocuCraft Pro**, a multi mode of extraction tool supporting No OCR, OCR, and AI enabled extraction.
- To gain practical knowledges of Agile methodologies, DevOps, and full-stack development.

1.5 Methodology

The development process followed an Agile (Scrum) framework, focusing on iterative sprints and team collaboration. The methodology relied on two types of data:

- **Primary Data:** This includes internal documentation from DataSoft, meeting notes with the Business Analyst team, and hands on data gathered during the testing and implementation of the projects.
- **Secondary Data:** Information gathered from online technical documentation (e.g., Flask, Redis, and Tesseract docs), research papers on OCR accuracy, and open-source GitHub repositories.

1.6 Literature Review

A review of existing tools shows that while global LMS and OCR platforms are powerful, they often lack the "lightweight" feel needed for small teams or the specific localization required for Bangla scripts. Research indicates that legacy encoding like Bijoy is a major obstacle in document mangement. Furthermore, traditional OCR engines like Tesseract are now being surpassed by multimodal GenAI models (like Gemini) in terms of contextual understanding and layout preservation, which is a key trend I incorporated into DocuCraft Pro.

Chapter 2

Company Profile

2.1 Company Overview

DataSoft Systems Bangladesh Limited was established in 1998 as a pioneer in the ICT industry of Bangladesh[1]. It holds the prestigious distinction of being the first software company in Bangladesh to be documented as a **CMMI Level 5** Software Development company. With over two decades of experience, DataSoft become a global solution provider and serving Fortune 500 companies and contributing on complex projects across three continents

As a CMMI Level 5 organization, the company follows world class standards in project management. This high level of maturity ensures highest quality of software delivery and continuous improvements to meet the evolving market demands.

2.2 Vision and Mission

Datasoft is focused to build a technologically advanced nation through innovation and exploration.

2.2.1 Vision

The vision of DataSoft is to deliver quality solutions to build **The Digital Delta** powered by innovation. They aim to set the standards in the global IT landscape.

2.2.2 Mission

The mission is to maintain the status of a leading software solution provider while ensuring the benefit of customers, shareholders, and employees. This is achieved by:

- Developing and exploiting the remarkable expertise and knowledge of their people.
- Developing a distinctive competence in process development.
- Ensuring customer satisfaction remains paramount to business success.

2.3 Operations and Strategic Partnerships

DataSoft operates with a global mindset and maintains strong ties with international organizations. A significant part of their strategy is their **collaboration with Japan**. Through projects like the JICA backed initiative for skill development, DataSoft has been a bridge between the Bangladeshi talent pool and the Japanese tech market.

The company works closely with the **Bangladesh Association of Software and Information Services (BASIS)** and has a long standing history of cooperation with the **Japan International Cooperation Agency (JICA)** to enhance the technical capabilities of local engineers.

2.4 Impactful Products and National Contributions

DataSoft is well known for handling nation building scale projects that touch millions of lives daily.

2.4.1 Rapid Pass (Metro Rail Integration)

One of the most crucial contributions is the implementation of **Rapid Pass**, the country's first *One Card for All Transport* solution. In collaboration with JICA and the Dhaka Transport Coordination Authority (DTCA), DataSoft build clearing house system which enables seamless ticketing for the Dhaka Metro Rail and other public transport services.

2.4.2 Financial and Microfinance Solutions

The company's flagship commercial product, **Microfin360** is used by thousands of microfinance institutions. It provides a comprehensive MIS and accounting platform that expanded to countries like Canada, India, and Nepal.

2.4.3 Infrastructure and Port Automation

DataSoft implemented the **Chittagong Port Automation (CTMS)**, which significantly improved the efficiency of the country's main maritime gateway. They were also involved in the national ID system, e-passport systems, and the Jute Genome Project that showcased their versatility.

2.5 Software Development Process and Technology Stack

DataSoft utilizes a hybrid development model that combines the rigorous controls of CMMI Level 5 with the flexibility of **Agile (Scrum and Kanban)** methodologies. This ensures timely delivery without compromising on documentation or security.

Technology Profile:

- **Frameworks & Languages:** Extensive use of Python (Flask/Django), Java (Spring Boot), .NET, and PHP (Laravel).
- **Frontend:** Modern stacks including React.js, Angular, and Vue.js.
- **Emerging Tech:** A strong portfolio in **IoT**, Artificial Intelligence (AI), Machine Learning (ML) and Blockchain.

2.6 Company Culture and Values

The culture at DataSoft is built on the belief that *Great teams build great companies*. There is a heavy emphasis on continuous training and mentorship, where senior developers guide interns to adapt to global market demands.

Core Values:

- **Integrity:** Being honest and fair in dealings with customers and partners.
- **Result-Oriented:** Seeking continuous improvement through aggressive but attainable goals.
- **Innovation:** A commitment to innovate because it transforms the way customers do business.

Chapter 3

Training and Development

3.1 Training and Hands-on Experience

The training process at DataSoft Systems Bangladesh Limited was mainly focused on hands-on experience within a professional workspace. Apart from academic projects, the training here pushed me to work with real world codebases and solve problems that created direct impacts on the company's output.

This practical approach shows how the theories of computer science concepts are applied to build scalable software. By working on projects like **Training PRO** and **DocuCraft Pro**, I gained a deep understanding of complex workflows of industry standards.

3.2 Skill Development at DataSoft

At Datasoft, I tried to improve my technical skills, The transition from basic programming to understand full stack systems enabled me to explore new new libraries and frameworks. This development timeline sharpen my ability to write clean code, maintain them under strict quality guidelines.

3.3 Team Collaboration and bKash AML Team

During my internship, I had the privilege to work within the **bKash AML Team**. Even though I was the most junior member of the team, I never felt out of place. Every member of the team was very helpful and they always praised my engineering mindset and spirit. The office culture was very inclusive and positive and they celebrates everyone's birthday together that made the professional journey much more engaging and enjoyable.

3.4 Mentorship and Professional Guidance

The mentorship and guidance I received from my team were truly immense. My team lead always guide me, help me to understand complex system architectures. When I struck in bugs or faced any issues regarding the projects, my colleagues directly helped me solve that, provide good resources for my learning. Also, they shared their real life experiences and gave advice on how to think like a real engineer and utilize the resources. There were times when I tried my best but did not get the desired result, I became very frustrated but my team never let my confidence drop. They motivated me highly and pushed me until the problem was solved.

3.5 Personal Growth and Professional Skills

The internship also contributed heavily to my personal growth and developing soft skills. I had the opportunity to give multiple presentations in front of a large audience which improved my public speaking and communication skills. It boosted my confidence and made me a more active and consistent performer in my professional life. Moreover, I learned the importance of time management by working with different project responsibilities.

Chapter 4

Proposed Methodology

4.1 Design Process or Methodology Overview

The development of both projects followed an **Agile (Scrum)** framework which required iterative progress then ongoing feedback and adaptability. Our team conducted daily stand up meetings to review updates, identify potential challenges and plan accordingly for the day. This approach played a crucial role in terms of collaboration and project management. Version control was handled using **Git**, while **GitHub** functioned as the main repository for source code supporting team oriented development, code reviews. The design process for each project was customized based on the unique goals and technical requirements:

- In the case of **Training PRO**, the development followed a full stack web application model. It involved backend REST API, implementing a persistent data storage layer and designing user interfaces using HTML and JavaScript.
- For **DocuCraft Pro**, the approach was to build a layered processing architecture that can handle different PDF types through three distinct extraction modules: No OCR, OCR, and GenAI-powered.

4.2 Project 1: Training PRO - A LMS Platform

The Training PRO platform is a modern and lightweight Learning Management System (LMS) that ensures asynchronous learning features.

4.2.1 Detailed Implementation

The platform is designed as a full stack web application. The backend is a RESTful API built with Python, on the other hand the frontend consists of HTML pages that ensures a clean, role based user experience.

The core of Training PRO involved with content management and progress tracking systems. This system is designed to support a wide range of educational media. The implementation details are as follows:

- **Role Based Architecture:** The system has different interfaces and functionalities for two primary user roles: **instructors** and **students**. Instructors are granted administrative privileges and they can create courses and manage content but students have access to the learning materials within their enrolled courses.
- **Multi-Format Content Delivery:** Instructors can upload content in five different formats: video (MP4), documents (PDF, DOCX), presentations (PPTX) and audio (MP3, WAV) . Each file type is handled by a backend service to store the file in an organized directory structure and records its metadata in a Redis database . One of the crucial feature is the automated processing of PowerPoint files. Where the backend converts each slide into a high resolution PNG image, enabling PPTX files to be viewed directly in the browser without requiring any plugins.
- **Granular Progress Tracking:** The platform's most sophisticated feature is real time progress tracking. For every student, the system stores their progress in Redis. In terms of videos and audio, progress is calculated as a percentage of the total duration watched. For PDFs and PPTX files, it is tracked by the highest page or the slide number reached. For DOCX, progress is determined by the user's scroll depth within the document viewer. This helps the students to seamlessly resume their learning from exactly where they left off.

4.2.2 Tools, Languages, and Frameworks Used

- **Python:** The primary backend programming language, chosen for its simplicity, extensive libraries and strong support for web development.
- **Flask:** It is a Lightweight Python microframework used to build the REST API. Its flexibility was ideal for creating the distinct services for content management.[6]
- **Redis:** A high performance, in memory key-value store used for managing user sessions and storing all real time progress tracking data.[5]
- **HTML5, CSS3, and JavaScript (ES6+):** These core web technologies were used to create the user interfaces, including the role based dashboards and the specialized media viewers.
- **python-pptx, Pillow, and pdf2image:** A suite of Python libraries used in the backend for processing uploaded files, specifically for converting PowerPoint slides and PDF pages into images.

4.2.3 Purpose of the Project

The core purpose of Training PRO was to develop a versatile solution. Within complex, monolithic systems, this project aimed to create a lightweight, scalable and easy to use platform for enhancing learning experience. Training PRO provides a holistic and engaging educational environment for both the instructor and learners.

4.2.4 Project Screenshots

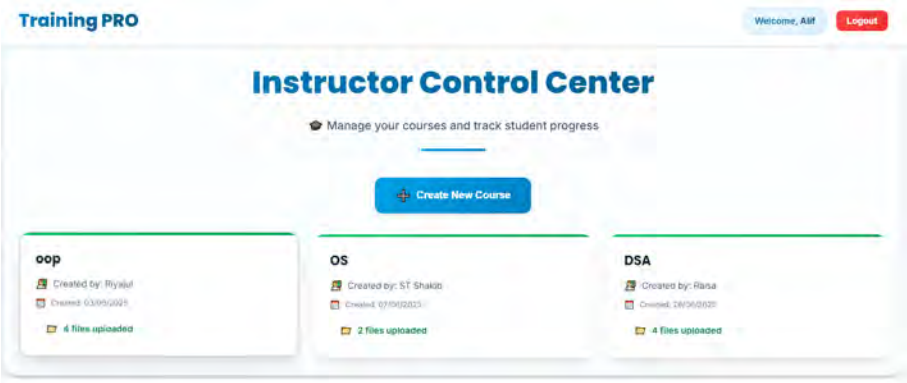


Figure 4.1: The Instructor main dashboard to manage courses.

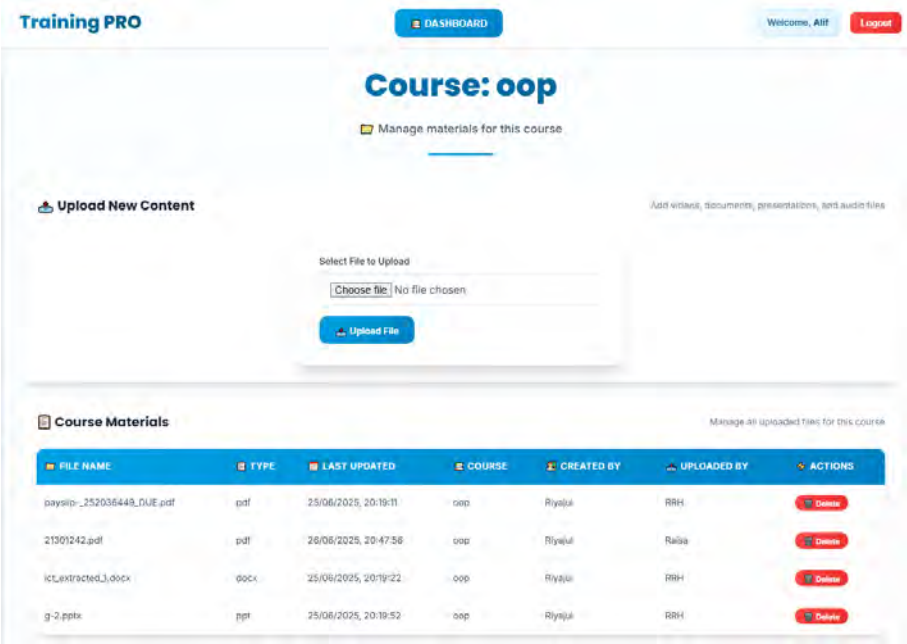


Figure 4.2: The course management interface for an individual course.

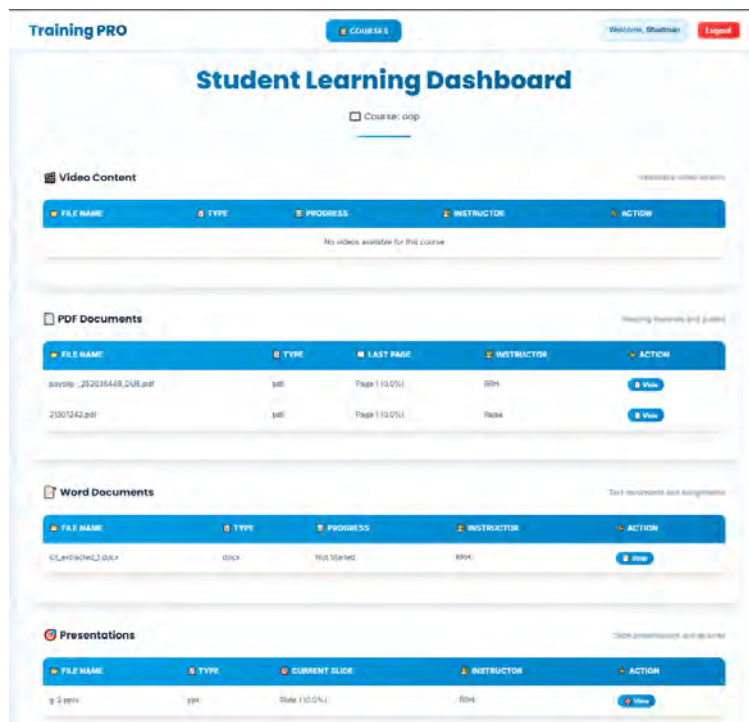


Figure 4.3: The Student Learning Dashboard interface in Training PRO.

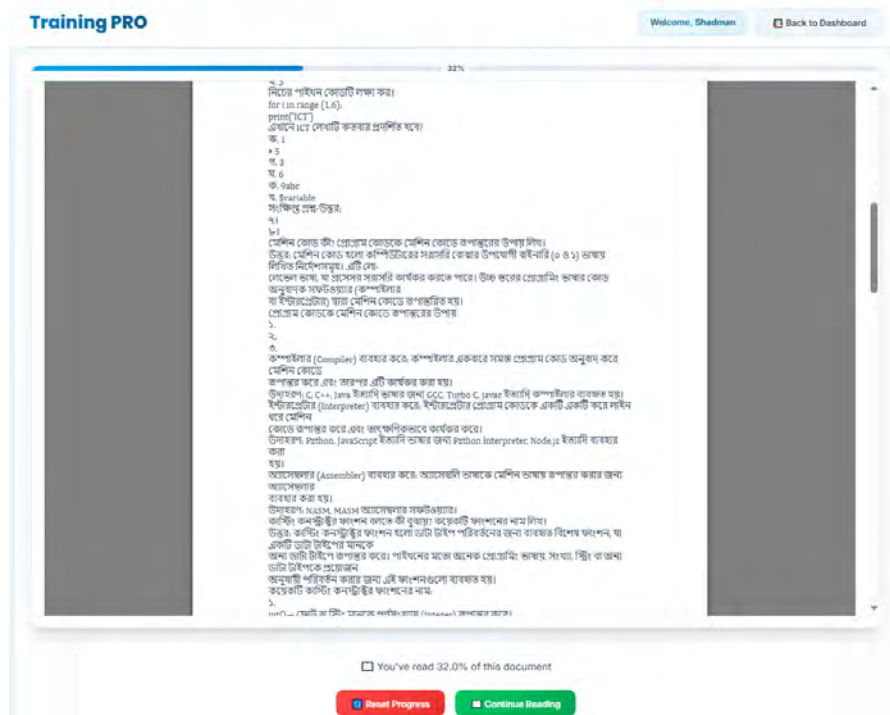


Figure 4.4: Example of Docx displaying the student's scroll-based progress (32.0%)

4.3 Project 2: DocuCraft Pro: PDF Text Extraction System

DocuCraft Pro is an advanced, multi method document processing tool developed to solve the complex challenges of extracting text from various types of PDFs, particularly focusing on Bangla script and legacy Bijoy encoding.

4.3.1 Detailed Implementation

The system is designed as a modular, multi stage processing pipeline that allows users to select the most appropriate extraction method based on their document's characteristics.

The workflow of the pipeline can be summarized as follows:

1. **File Upload and Method Selection:** The process begins when a user uploads a PDF file through the responsive web interface. The Flask backend validates the file extension and ensures that the size is within the 16MB limit. At this stage, the user also chooses one of three processing methods like No OCR, OCR, or GenAI and selects the desired output format (TXT or DOCX).
2. **Pre-processing and Language Detection:** At first the system sanitizes the filename and then initiates a language detection check. This step is crucial for identifying whether the content is English or Bangla, specifically if the Bangla text is using Bijoy encoding system then it goes for special conversion layer.
3. **Multi-Method Text Extraction:** This is the core state where the chosen algorithm is executed:
 - **No OCR:** Will performs direct text extraction using libraries like PyMuPDF and PyPDF2 then applying Bijoy-to-Unicode conversion[2] if necessary.
 - **OCR Method:** Converts each page into a high resolution 300 DPI image and then uses the Tesseract OCR[7] for extraction. .
 - **GenAI Method:** In terms of complex layouts, images are sent to the Google Gemini API[4], which uses multimodal vision models to extract text with high accuracy.
4. **Formatting and Document Assembly:** After the text is successfully extracted, the `python-docx` library are used to assemble the content into a structured format. This ensures that line breaks, paragraph structures and page breaks are preserved as closely as possible like the original document.
5. **Success Feedback and Download:** The system provides real time status updates through a processing overlay. Once the conversion is complete, a success animation is displayed and the user is provided with a direct download link for the final editable document.

4.3.2 Project Screenshots

The screenshot displays the DocuCraft PRO web interface. At the top, a blue header contains the text "DocuCraft PRO" and "Crafting perfect text from documents". Below this is a large white area with a blue cloud icon and an upward arrow, followed by the text "Drag & Drop your PDF here" and "or". A "Browse File" button is positioned below the text. The "Select Processing Method:" section follows, with three radio button options: "No OCR" (selected), "OCR Based", and "GenAI Based". Each option includes a brief description. The "Output Format:" section has two radio button options: "TXT" (selected) and "DOCX". A "Process PDF" button with a gear icon is located below the output format section. The footer is a blue bar with "DocuCraft Pro" and "© 2025 Datasoft. All Rights Reserved."

DocuCraft PRO
Crafting perfect text from documents

Drag & Drop your PDF here
or
[Browse File](#)

Select Processing Method:

☒ **No OCR** - Direct text extraction with auto language detection
Fast and accurate for text-based PDFs

☐ **OCR Based** - Image-based extraction with auto language detection
Best for scanned or image-based documents

☐ **GenAI Based** - Google Gemini AI-powered extraction
High accuracy for complex layouts and mixed languages (Requires API key)

Output Format:

☒ **TXT** ☐ **DOCX** Choose between TXT or DOCX output formats

[Process PDF](#)

DocuCraft Pro
© 2025 Datasoft. All Rights Reserved.

Figure 4.5: The main interface for DocuCraft PRO.

Chapter 5

Challenges and Learning Outcomes

5.1 Technical Challenges

Developing complex systems like **Training PRO** and **DocuCraft Pro** within the environment of DataSoft Systems Bangladesh Limited helps to improve real world problem solving skills and overcome the technical hurdles through continuous efforts.

5.1.1 Challenges in Training PRO

The primary challenge during the development of Training PRO was the implementation of a realtime progress tracking system for various media formats.

- **Working with Redis:** This was my first time working with Redis for session management and data storage. Understanding how redis works and storing different media types in a key-value pair was a bit challenging.
- **Complexity of PPTX Files:** Handling Pdf and Docx files was a straightforward approach but the workflow of Pptx files was complex. Converting individual slides into high resolution images and maintaining layout integrity was difficult. I had to explore a lot and go through trial and error to reach this solution.

5.1.2 Challenges in DocuCraft Pro

Building an intelligent document processor for the local context introduced unique obstacles, particularly when dealing with non standard document formats.

- **Bijoy to Unicode Conversion:** Converting legacy Bijoy encoding into standard Unicode was one of the most difficult tasks. I searched for several github repositories and tested each of them and then fine tune to integrate to the workflow.
- **OCR Model Performance:** I explored several models like **Tesseract**, **PadleOCR**, and **EasyOCR**. Since I did not have access to a dedicated GPU at the company, I use Kaggle for testing. Unfortunately, the performance and accuracy of some of these models were not satisfactory.

- **GenAI and API Limits:** Moving to a GenAI approach with **Google Gemini** was a great decision for accuracy but it introduced new challenges. The API tier limits often caused failures when processing documents with a large number of pages.

5.2 Technical Learning and Professional Growth

Despite these challenges, the internship provided an incredible learning opportunity to explore various domains of software engineering.

5.2.1 Full-Stack Development Mastery

Working on **Training PRO** required me to deep dive into the **Flask** framework. I gained deep knowledge of **Redis** and became familiar with many new libraries, which eventually sharpened my fullstack development skills. I learned how to build a scalable backend that could handle multiple content efficiently.

5.2.2 AI and Document Processing Expertise

Through **DocuCraft Pro**, I explored different OCR engines like **Tesseract**, **PaddleOCR** and **EasyOCR** to understand their strengths and weaknesses. I also mastered **prompt engineering** to work with GenAI based approaches. This taught me how to work with AI APIs, understand their limits and optimize the system architectures for better outcomes.

5.2.3 Experience with AI Chatbots and SQA

I had the privilege to work with the **NILA chatbot** which was an AI-powered solution for **DPDC**. During that time, I collaborated closely with the testing team, which helps me to gain valuable knowledge of **Software Quality Assurance (SQA)**.

5.2.4 Android Development Exploration

I also spent time working with the **Android team** at DataSoft. In this domain, I explored **Java** and the **Android SDK**. I explored **map integration** for various mobile projects which broadened my understanding of the mobile ecosystem development.

5.2.5 DevOps and Infrastructure

In the second month of my internship, I had the opportunity to explore the **DevOps** side of software engineering. I came to know about **Docker** for containerization and explored **CI/CD pipelines**. I also worked with load balancing using **Nginx** in one of the projects that gave me insights on how to ensure system availability and performance.

5.3 Summary of Learning Outcomes

I worked without local GPU support and go through Gemini API limits, it taught me how to be resourceful and adaptable. From working in cloud based environments, learning optimizing prompts helps me to think like a professional engineer. Being part of a **CMMI Level 5** company like DataSoft taught me the value of **Agile** practices and the importance team collaboration.

Chapter 6

Conclusion

6.1 Summary of Findings

The internship at DataSoft Systems Bangladesh Limited was a very significant experience in terms of technical and professional growth. During this time, I successfully collaborated and developed two types of software solutions: **Training PRO** and **DocuCraft Pro**.

For **Training PRO**, it was a lightweight, asynchronous LMS that can manage various types of contents with progress tracking. The use of **Redis** makes the system highly efficient to handle the real time progress easily.

For **DocuCraft Pro**, a multi method approach is essential. While direct extraction and OCR method works well with simple files, the **GenAI** method helps to overcome the formatting issues and complex challenges of mixed language documents.

6.2 Professional Contributions

My time at Datasoft was not only involved with SDLC and coding tasks, I actively participated and collaborate with the team in the professional ecosystem of the company.

6.2.1 Collaboration with Business Analysts

While working with the Business Analyst team, I used to share my views and try to provide creative ideas during team meetings and project discussion. It helps me to contribute to the projects from both technical and business perspective. My supervisors appreciated my spirit and engineering mindset.

6.2.2 Technical Presentations

I had the opportunity to give presentations on **NILA chatbot** (AI powered chatbot of DPDC) that creates a positive impression of my team within the company.

6.2.3 Team Integration

I participate in **scrum meetings**, share the progress, collaborate with my colleagues to ensure smooth project workflows. Moreover, I participated in various cultural events and understood the importance of healthy workplace culture.

6.3 Recommendations for Future Work

Based on the current state of the prototypes and feedback from the team at DataSoft, there are several key areas for future development:

6.3.1 Future Plans for Training PRO

- **Zoom API Integration:** To create and schedule meetings seamlessly directly from the Training PRO dashboard, the team plans to integrate Zoom with the help of API.
- **Interactive Assessment Engine:** Instructors can automatically create MCQ type questions from the uploaded contents and take quizzes and evaluate the performance within the LMS.

6.3.2 Future Plans for DocuCraft Pro

- **Local Open-Source OCR Models:** Gemini API has limits and face challenges processing large documents. Open source models like **DeepSeek OCR** or **Chandra OCR**[3] can help to solve the obstacle.
- **Edge Device Optimization:** PDF contains sensitive information and working with locally on edge devices will ensure that processing happens on the local machine without uploading data to external cloud servers. It will enhance the overall security and privacy.

6.4 Concluding Remarks

This internship has been a vital turning point in my educational journey. Working at a **CMMI Level 5** company like DataSoft provided me with a deep understanding of industry standard practices and professional responsibility. I became a confident contributor by overcoming the technical difficulties and collaborate with my team. I am very grateful to the Bkash AML team at Datasoft for their mentorship and giving me the opportunity to work with these innovative projects.

Bibliography

- [1] DataSoft Systems Bangladesh Limited. “Datasoft: The first cmmi level 5 software development company in bangladesh. ”[Online]. Available: <http://datasoft-bd.com/>. (accessed: 05.02.2026).
- [2] A. Mehady. “Bijoy-to-unicode-file-converter: Python tool for batch converting legacy bijoy files. ”[Online]. Available: <https://github.com/almehady/Bijoy-to-Unicode-File-Converter>. (accessed: 05.02.2026).
- [3] Datalab. “Chandra: Highly accurate open-source ocr model for structured document understanding. ”[Online]. Available: <https://github.com/datalab-to/chandra>. (accessed: 05.02.2026).
- [4] Google. “Gemini api overview. ”[Online]. Available: <https://ai.google.dev/models/gemini>. (accessed: 05.02.2026).
- [5] R. Ltd. “Redis documentation. ”[Online]. Available: <https://redis.io/docs/>. (accessed: 05.02.2026).
- [6] P. Projects. “Flask documentation. ”[Online]. Available: <https://flask.palletsprojects.com/>. (accessed: 05.02.2026).
- [7] R. Smith. “Tesseract ocr. ”[Online]. Available: <https://github.com/tesseract-ocr/tesseract>. (accessed: 05.02.2026).