 **Al Maaref University Faculty of Sciences**

**Department of Management Information System**

Summer 2023 - 2024

**Practical Training (MIT497) Report**

**Student ID:** 10121623

**Student Name:**  Zahraa Ahmad Neameh

**Internship Title:** Data Science & Engineering Internship

**Company Name:** dgPad

**Location:** Beirut, Lebanon

**Company Coordinator:** Mr. Kassem Shehady

**University Coordinator:** Dr. Kassem Danach

**Table of Contents**

[Company Information 3](#_Toc4748)

[Company Products and Services 3](#_Toc4749)

[Internship Information 3](#_Toc4750)

[Internship Abstract 4](#_Toc4751)

[Internship Plan and Responsibilities 4](#_Toc4752)

[Achievement Details 5](#_Toc4753)

[Achieved Development Areas 5](#_Toc4754)

[Hardware Environment 5](#_Toc4755)

[Learned Technologies 6](#_Toc4757)

[Used Tools 6](#_Toc4758)

[Product Benefits 6](#_Toc4759)

[Product Repository Link 7](#_Toc4760)

[Critical Analysis 8](#_Toc4762)

[Challenges Faced 8](#_Toc4763)

[Compatibility of University Courses 8](#_Toc4764)

[Personal Benefits 8](#_Toc4765)

[Professional Benefits 9](#_Toc4766)

[Comments on the Company 9](#_Toc4767)

[Work Opportunity 10](#_Toc4768)

# Company Information

**Company Name:** dgPad

**Company Domain:** Technology, Information, and Internet

**Company Size:** 2-10 employees

**Company Sales Size (if available):** N/A

**Company Website:** [dgpad.io](https://dgpad.io/)

## Company Products and Services

* .Net Full Stack Engineering
* Java Engineering
* Data Science & Engineering



# Internship Information

|  |  |
| --- | --- |
| **Training Title** | **:** Data Science & Engineering |
| **Company Department** | **:** Technology, information, and Internet |
| **Internship Duration (hours)** | **: 56** hours |
| **Company Coordinator** | **:** Mr. Kassem Shady |
| **Company Coordinator (email)** | **:** Kassem.Shady@gmail.com |
| **University Advisor** | **:** Kassem.Danach@mu.edu.lb |
| **University Advisor (email)** | **:** Kassem.Danach@mu.edu.lb |

## Internship Abstract

## During my internship, I focused on advanced data analysis and API development to derive actionable insights from complex datasets. This involved applying data science techniques such as sentiment analysis, entity recognition, and trend analysis to process collected textual data. My responsibilities included developing and enhancing a Flask API to support these advanced analytical functions, storing processed results in MongoDB, and creating an interactive data visualization dashboard. The global objective of this training was to build a comprehensive data-driven application, covering end-to-end processes from data collection and processing to visualization and analysis. This experience equipped me with practical skills in data science, full-stack data engineering, and API development for real-world applications. Internship Plan and Responsibilities

**Internship Plan and Responsibilities**

1. **Data Collection and Initial Processing**
   * Collected and preprocessed datasets, ensuring data quality and consistency.
   * Cleaned and formatted textual data to prepare it for further analysis.
2. **Data Visualization and API Setup**
   * Created visualizations with am Charts to represent initial data insights.
   * Set up a Flask API to serve data and provide basic query functionalities.
   * Integrated MongoDB as a backend database for efficient storage and retrieval.
3. **Advanced Data Analysis**
   * Performed sentiment analysis on collected text data using NLP libraries (e.g., Text Blob, NLTK).
   * Stored sentiment results in MongoDB and created endpoints to support sentiment-based queries.
   * Conducted entity recognition to extract named entities (people, places, organizations) from articles.
     + Enhanced the API to support entity-based queries and filtering.
     + Stored extracted entities in MongoDB for efficient querying.
4. **Trend Analysis**
   * Analyzed trends in sentiment and keyword usage over time.
   * Visualized trends through line and bar charts, showing topic evolution and sentiment changes.
5. **API Enhancements for Advanced Queries**
   * Developed endpoints for retrieving most positive/negative articles, entity-based article retrieval, and trend analysis.
   * Added endpoints to query and visualize trends based on specific keywords or entities.
6. **Comprehensive Dashboard Creation**
   * Built an interactive dashboard that combines all visualizations and insights.
   * Enabled filtering options for sentiment, entities, and date ranges, enhancing user interaction.
7. **Security, Optimization, and Deployment**
   * Implemented security measures for the API, including rate limiting and API key validation.
   * Optimized MongoDB queries and API performance to handle large datasets.
   * Deployed the Flask application and dashboard on a cloud server, making it accessible to end-users.

# Achievement Details

## Achieved Development Areas

* **Backend**: Developed and optimized a Flask API to serve advanced data insights, enabling functionalities like sentiment analysis, entity recognition, and trend analysis.
* **Frontend**: Designed an interactive data visualization dashboard for users to explore insights through filtering and trend analysis.
* **Database**: Used MongoDB to store and retrieve processed data, including sentiment scores and extracted entities, ensuring efficient querying and scalability.

## Hardware Environment

 **Development Computer**: Intel Core i7, 16GB RAM, 512GB SSD – used for development and testing.

 **Server Environment**: Cloud-based server with 4vCPU, 8GB RAM, and 100GB storage for hosting Flask API and MongoDB database, ensuring smooth performance during data processing and API calls.

* **Operating Systems**: Primarily worked on Windows for development and testing, with deployment on a Linux-based server for the production environment.

## Learned Technologies

1. **Flask**: A lightweight web framework used for creating the API to serve data insights.
2. **MongoDB**: A NoSQL database utilized for flexible and efficient storage of data, supporting complex queries for advanced analytics.
3. **Text Blob**: A Python library for natural language processing used for sentiment analysis.
4. **spaCy**: An NLP library employed for entity recognition, extracting relevant entities like people and organizations from text data.

## am Charts: A charting library used to create interactive data visualizations in the dashboard, allowing users to track trends over time.

## Used Tools

1. **PyCharm**: An integrated development environment (IDE) used for coding and debugging the Python application.
2. **Postman**: API testing tool used to test and verify Flask API endpoints during development.
3. **Git**: Version control system for tracking changes, collaborating on code, and managing different versions.

## Product Benefits

 **Resulting Product**: The product is a data-driven application with an API and dashboard that enables advanced data analysis. This can benefit companies or organizations needing in-depth insights from textual data, such as sentiment trends, key entities, and overall data evolution.

##  ****Current Use****: This application can be used by businesses, research teams, or media agencies for data analysis. It is currently in a prototype phase and is potentially deployable for real-world applications. Product Repository Link

## Product Repository Link

**Product Repository on GitHub** [*https://github.com/zahraaneameh2019/mybootcamp.git*](https://github.com/zahraaneameh2019/mybootcamp.git)

# Critical Analysis

## Challenges Faced

During my internship, I encountered several challenges:

1. **Complex Data Processing**: Working with large, unstructured datasets required advanced preprocessing and careful handling to maintain data quality. This required extra time and effort to implement efficient data cleaning and formatting processes.
2. **API and Database Optimization**: Integrating MongoDB with the Flask API posed some challenges, especially when handling large datasets. Query optimization and API performance tuning became essential to ensure smooth functionality.
3. **Advanced NLP Techniques**: Applying sentiment analysis and entity recognition at scale required a deep understanding of natural language processing libraries, which I had to learn and implement quickly.
4. **Deployment and Security**: Implementing security measures, such as rate limiting and API keys, was challenging as I had to balance user accessibility with data protection.

## Compatibility of University Courses

This internship provided a valuable practical application of concepts learned in my university courses, especially in programming, data management, and analytics. While my studies covered data structures, database design, and introductory data science, the internship expanded on these areas significantly by introducing advanced NLP techniques, API development, and real-world deployment practices. The combination of theoretical knowledge from class and practical challenges from the internship provided a complete learning experience that complemented my coursework well.

## Personal Benefits

This internship was highly beneficial on a personal level:

* **Mindset**: I developed a mindset focused on problem-solving and adaptability, particularly when handling unexpected challenges.
* **Confidence**: Gaining hands-on experience with complex technologies like Flask, MongoDB, and NLP libraries boosted my confidence in my technical abilities.
* **Networking and Collaboration**: Working alongside professionals allowed me to improve my communication skills and learn from others, enhancing my ability to collaborate in a team environment.

## Professional Benefits

On a professional level, the internship was instrumental in developing my technical and project management skills:

* **Technical Skill Enhancement**: I gained expertise in full-stack data engineering, API development, and advanced data analysis, which are in high demand in today’s job market.
* **Portfolio Building**: The resulting application, with a repository link and interactive dashboard, serves as a valuable addition to my professional portfolio, showcasing my capabilities to potential employers.
* **Career Direction**: This internship solidified my interest in data science and software engineering, helping me focus my career path towards data-driven application development.

## Comments on the Company

The company provided a supportive and resource-rich environment for learning, with access to experienced mentors and up-to-date technologies.

* **Pros**: The team was collaborative, and I had the freedom to experiment with various tools and libraries. The company also emphasized the importance of best practices in coding and security, which improved my approach to software development.
* **Cons**: At times, I encountered limited documentation for some of the company-specific tools, which required additional effort to understand and integrate them into my work.

Overall, the training at this company was worth it. I gained hands-on experience with modern technologies and data analysis techniques that would have been difficult to acquire in a purely academic setting.

## Work Opportunity

I am open to continuing work with this company, provided there are opportunities aligned with my interests in data science and software development. The practical experience gained through this internship, along with the company's support, has made it an appealing place for potential full-time employment.