|  |  |
| --- | --- |
| Welcome | **Al Maaref University**  **Faculty of Business Administration**  **Department of MIT**  Summer 2024 - 2025 |

**Practical Training (MIT 497) Report**

**Student ID :** 10121673

**Student Name :** Haydar Ramadan Mohamad

**DataScience Bootcamp (Python)**

DgPad-Mentis

Beirut, Lebanon

Company Coordinator : Kassem Shehady

University Advisor : Dr. Rabih Tarraf

Table of Contents

[Company Information 3](#_Toc180751429)

[Company Products and Services 3](#_Toc180751430)

[Internship Information 4](#_Toc180751431)

[Internship Abstract 4](#_Toc180751432)

[Internship Plan and Responsibilities 4](#_Toc180751433)

[Achievement Details 5](#_Toc180751434)

[Achieved Development Areas 5](#_Toc180751435)

[Hardware Environment 5](#_Toc180751436)

[Software Platform 5](#_Toc180751437)

[Learned Technologies 6](#_Toc180751438)

[Used Tools 6](#_Toc180751439)

[Product Benefits 6](#_Toc180751440)

[Product Repository Link 6](#_Toc180751441)

[Product Images Overview 7](#_Toc180751442)

[Critical Analysis 11](#_Toc180751443)

[Challenges Faced 11](#_Toc180751444)

[Compatibility of University Courses 11](#_Toc180751445)

[Personal Benefits 11](#_Toc180751446)

[Professional Benefits 12](#_Toc180751447)

[Comments on the Company 12](#_Toc180751448)

[Work Opportunity 13](#_Toc180751449)

# Company Information

**Company Name :** DGPad-Mentis

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

**Company Size :** 2-10 employees

**Company Sales Size (if available) :** Not Available

**Company Website :** http://dgpad.io

## Company Products and Services

DGPad-Mentis specializes in offering training programs in the field of technology, focusing on helping fresh graduate students acquire skills to get their first job through a tech bootcamp with domain experts. The company aims to bridge the gap between theoretical learning and practical industry requirements by providing immersive bootcamps that equip students with real-world, industry-relevant skills. These include:

* Data Engineer Bootcamp
* .Net full stack engineer Bootcamp
* Java Engineer Bootcamp

# Internship Information

**Training Title :** Data Science Bootcamp Using Python

**Company Department :** IT

**Internship Duration (hours) :** 2 Months

**Company Coordinator :** Kassem Shehady

**Company Coordinator (email) :** kassemshehady@gmail.com

**University Advisor :** Dr. Rabih Tarraf

**University Advisor (email) :** rabih.tarraf@mu.edu.lb

## Internship Abstract

During the Data Science Bootcamp at DGPad-Mentis, the main focus was web scraping, data extraction, and advanced analytics using Python. I successfully extracted data from XML sources (Al Mayadeen Website), stored the articles as JSON files, and then managed this data within a MongoDB database.

I developed Flask APIs to allow seamless interaction with the stored data, and created visualizations using amCharts, which I integrated into a dynamic, data-driven dashboard. Additionally, I performed advanced data analysis, including sentiment analysis, named entity extraction, and trend analysis using a variety of natural language processing (NLP) tools such as spaCy, TextBlob, CAMeL, and Farasa.

These tasks contributed to building a comprehensive dashboard that offers the company actionable insights from news data, enabling more informed decisions. The internship not only allowed me to strengthen my Python and data science skills but also provided significant value to the company by enhancing their ability to work with large-scale data

## Internship Plan and Responsibilities

List of responsibilities and tasks assigned to you during your training and **that you successfully achieved, organized in chronological order**

1. Web Scraping with Python
   1. Parsed the website’s sitemap to locate articles and efficiently extract data.
   2. Scraped a large volume of news articles (10000 articles) from the Al Mayadeen website.
   3. Stored the scraped articles in a structured JSON files (about 9 files) for further processing.
2. Data Storage in Mongodb and Flask API Development
   1. Designed and implemented data storage solutions using MongoDB to efficiently manage the scraped data.
   2. Developed RESTful APIs (30 endpoints) using Flask to interact with the stored data, enabling access to the data for further analysis and visualizations.
3. Data Visualization Using AmCharts
   1. Created interactive and insightful data visualizations using amCharts, which provided key metrics and trends from the data.
4. Advanced Data Analysis and Insights
   1. Conducted **Sentiment Analysis** on the articles to gauge public sentiment on various topics.
   2. Performed **Named Entity Recognition** to extract and categorize key entities (people, organizations, locations).
   3. Conducted **Trend Analysis** to identify patterns in article content over time.
   4. Enhanced the APIs to deliver better results and improved functionality.
   5. Developed a comprehensive **dashboard** that combined all these features into a single platform for easy access to the data and insights.
5. Dashboard UI/ UX Design improvements

# Achievement Details

## Achieved Development Areas

* Backend**:** Developed APIs using Flask for managing and interacting with data stored in MongoDB. Implemented data analysis features such as sentiment analysis and entity extraction.
* Frontend: Created data visualizations using amCharts and built an advanced dashboard to display the results of data analysis.
* Database: Used MongoDB to store, and manipulate JSON files that were saved from web-scraped data.
* Data Analysis**:** Did advanced analysis on the data using libraries like spaCy, TextBlob, CAMeL, and Farasa for sentiment analysis, entity extraction, and trend analysis

## Hardware Environment

* Hardware Specs: Standard laptop with regular processing capabilities and internet connection.

## Software Platform

* Operating System: Worked on Windows, but most of the tools and libraries used can be used on whatever platform.
* Development Platform: Used PyCharm for coding, developing, and testing the scripts
* Database Platform: Used Mongodb to store and manage the JSON files.

## Learned Technologies

* Python: Programming Language
* Flask: web framework used for building APIs
* Mongodb: NoSQL database used to store and manage data
* AmCharts: JavaScript library for creating data visualizations
* SpaCy, Camel tools, TextBlob, farasa…: Libraries used for advanced data analysis
* Web Scraping Tools: Tools used for extracting data from XML websites (requests, Bs4…)

## Used Tools

* PyCharm: Integrated Development Environment (IDE) for Python development, including web scraping, API development, and backend work.
* Mongodb Compass: A graphical interface for querying and managing MongoDB collections.
* Postman: Tool for testing APIs to ensure their functionality.
* Github: Version control and collaboration platform where the code and projects were managed.

## Product Benefits

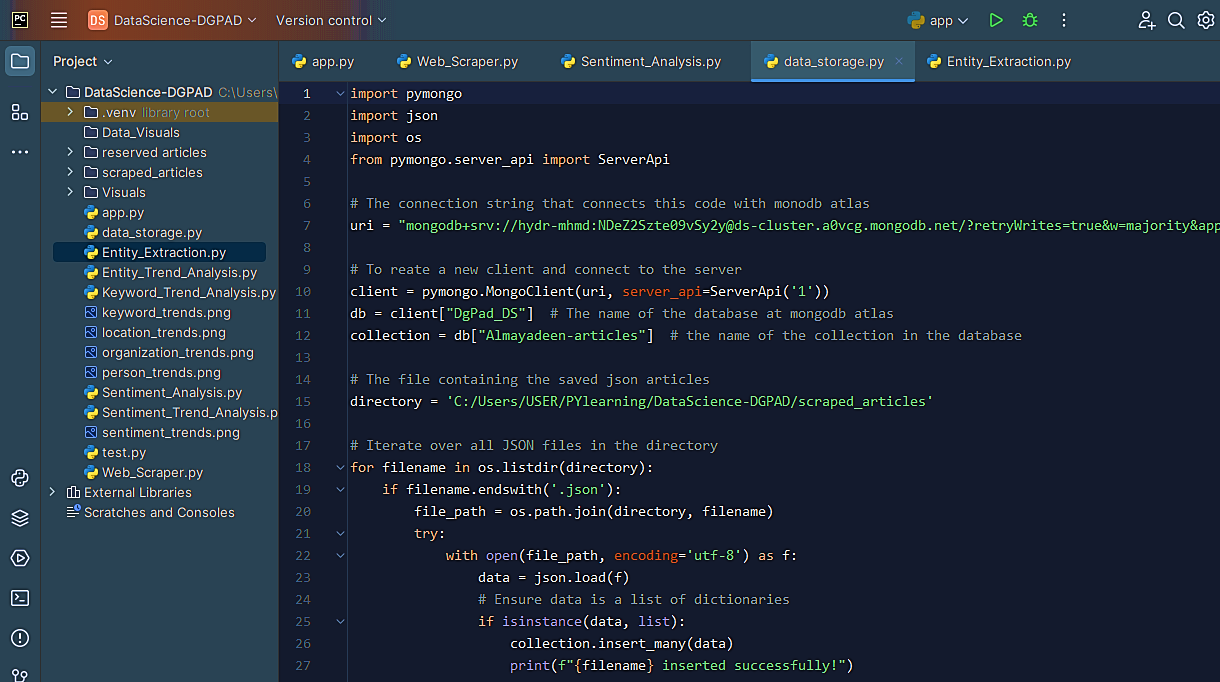
* The product resulting from my training is an **advanced data analysis dashboard** that provides insights into web-scraped news articles. It supports sentiment analysis, named entity recognition, and trend analysis, which makes it valuable for media companies, data analysts, and organizations interested in monitoring trends in the news.
* This is not currently being used commercially, but it was a learning project during my bootcamp and has the potential for real-world applications.

## Product Repository Link

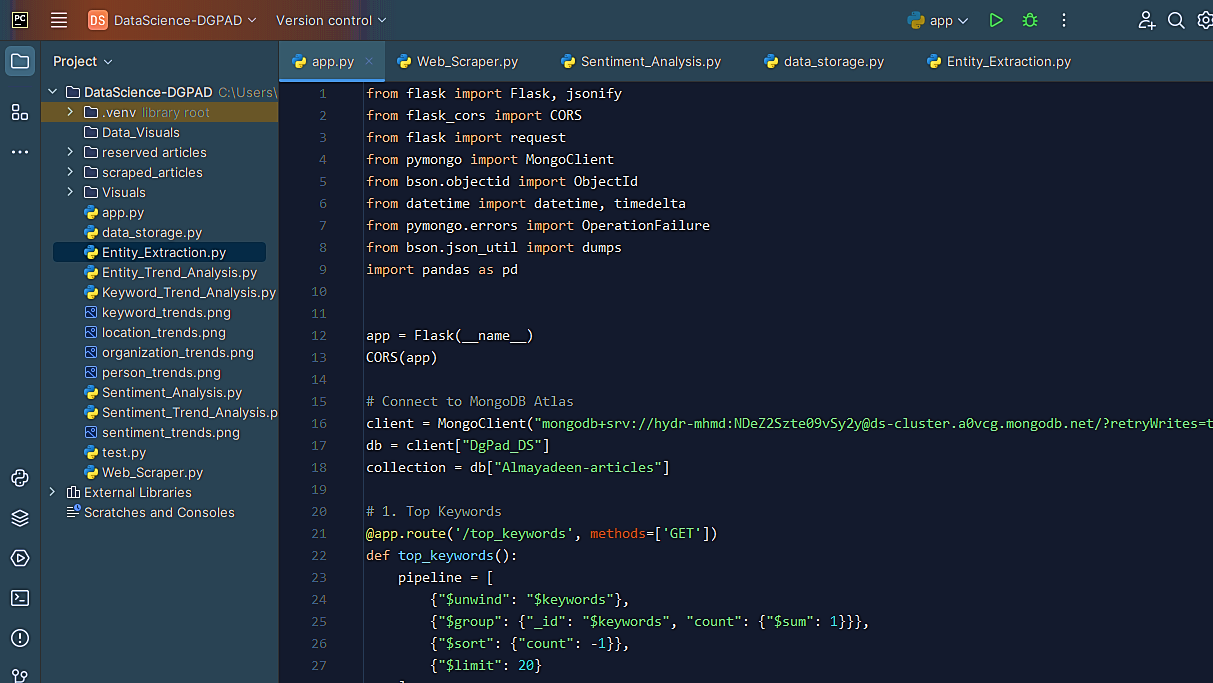
* Github Repository Link: https://github.com/HydrMhmad731/DgPad-DataSience-Project

## Product Images Overview

* 1st pic: This is a picture for the web-scrapping code done in PyCharm IDE using python as the programming language and libraries within (requests, bs4…), to scrape articles from Al Mayadeen Website.



* 2nd pic: This is the picture for data storage in MongoDb code, using the pymongo library to store the fetched articles.



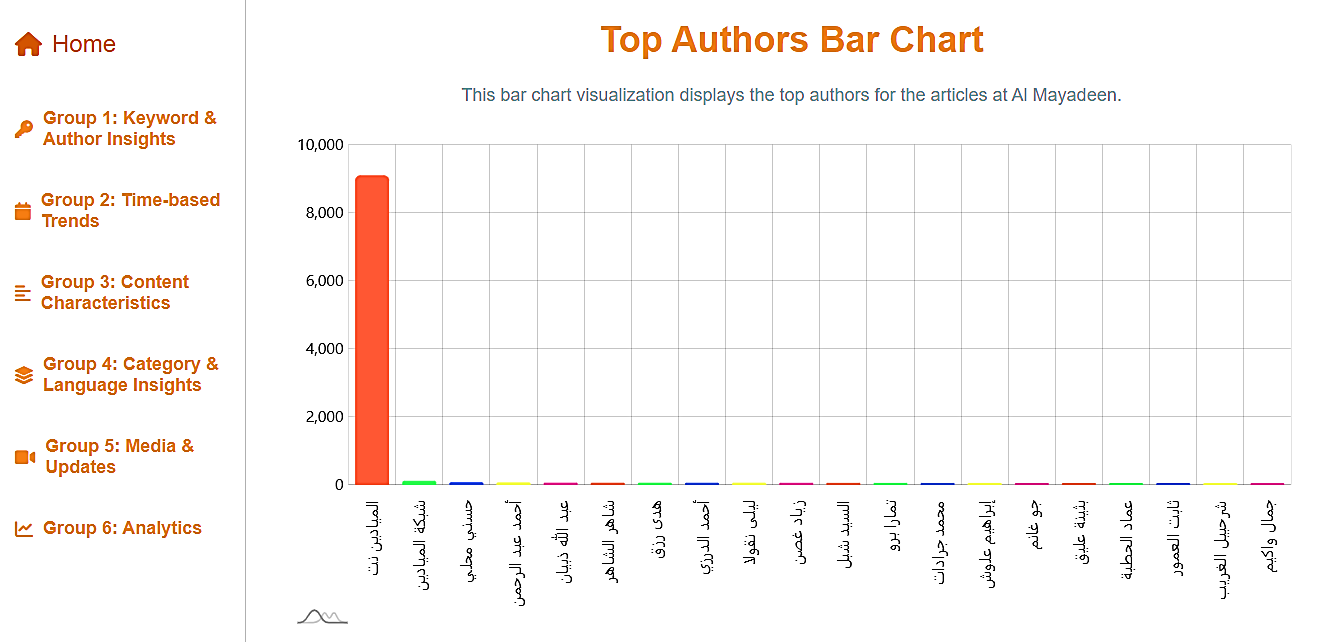
* 3rd pic: A picture of the flask code for creating APIs using the flask library with integration with pymongo to connect with the data inside mongodb.



* 4th pic: A picture inside mongodb showing the database, collection, and the stored articles.



* 5th pic: A picture of data visualization’s dashboard, where it is showing the top keywords word-cloud visualization, done by using AmCharts with integration with front end web development.



* 6th pic: Another picture of the dashboard , showing the top authors bar-chart visualization.

# Critical Analysis

## Challenges Faced

During the internship, I encountered several challenges:

* Learning New Libraries: Working with libraries such as **spaCy**, **CAMeL Tools**, and **Farasa** required time to understand and properly implement them for tasks like entity extraction and sentiment analysis.
* Web Scraping Difficulties: Parsing and scraping data from news websites was more challenging than anticipated, especially with handling website restrictions, varying XML structures, and ensuring clean, accurate data extraction.
* API development and Integration: Building efficient APIs using Flask and ensuring smooth integration with MongoDB required debugging and testing to ensure proper functionality, especially when handling large datasets
* Advanced Visualizations: Understanding how to use **amCharts** for dynamic visualizations was difficult at first, particularly when dealing with interactive features and customization.

I overcame these challenges by seeking help from documentation, online forums, and the support of my coordinator. These experiences improved my problem-solving skills and technical capabilities.

## Compatibility of University Courses

The internship aligned well with the theoretical knowledge I gained in university, particularly in courses related to:

* Data-Base Systems: My university course on databases helped me grasp MongoDB and understand how to efficiently manage and query large datasets.
* Programming and Algorithms: My Python programming class, and web development classes gave me the foundational skills necessary to set the ground for web scraping, creating visualizations, creating a well built dashboard and advanced tasks.

However, the practical, hands-on nature of the internship pushed me beyond classroom theory, especially in areas like **web scraping**, **building APIs**, and **data visualization**, which were not as deeply covered in the curriculum.

## Personal Benefits

On a personal level, this internship has been transformative in several ways:

* Confidence: I gained confidence in my technical skills, particularly in problem-solving and working with real-world data.
* Adaptability: Facing various challenges, such as learning new libraries and debugging code, has made me more adaptable and resilient when tackling unfamiliar problems.
* Mindset: My approach to projects has evolved. I now understand the importance of transforming tasks into smaller chunks, manageable steps and approaching them methodically.
* Collaboration: Interacting with my mentors and team members improved my communication skills and gave me valuable experience in a professional setting.

## Professional Benefits

Professionally, this internship has provided a significant boost to my career development:

* Technical Skills: I have deepened my knowledge of **Python**, **MongoDB**, and **Flask** for backend development. I also gained hands-on experience with **data analysis** and **visualization tools**.
* Portfolio Development: The project I worked on—building a data analysis dashboard—serves as an impressive addition to my projects, showcasing my ability to handle advanced projects.
* Industry Relevance: I learned how the skills I developed in data science are applicable in the real world, particularly for companies looking to derive insights from large datasets.
* Professional Network: Working with professionals in the industry, especially my coordinator, has expanded my professional network and provided me with valuable mentorship.

## Comments on the Company

* Pros
  + Mentorship: I received constant support from my coordinator, who provided valuable feedback and guidance throughout the project.
  + Work Environment: The company fostered a collaborative and encouraging environment, where I was given the freedom to explore different approaches to solve problems.
  + Learning Opportunity: I had the opportunity to work with cutting-edge tools and technologies that are highly relevant in today’s data science field.
* Cons
  + Project Scope: The scope of the project could have been more clearly defined at the beginning, which led to some confusion and delays during the development phase.

Overall, my experience at **DGPad-Mentis** was highly positive, and I believe the training was worth it, as it provided me with real-world experience and valuable skills that will help me in my career.

## Work Opportunity

At the moment, there hasn’t been a concrete discussion about continuing with the company after my internship. However, given the skills I’ve gained and the positive work environment, I would be open to future opportunities with **DGPad-Mentis**, should they arise.