# Data Scientist Work Report

Bayer AG

Monheim am Rhein

16.10.23 - 31.03.24

Maryana Smirnova 22109911

AIN-B-5

March 2024

# Description of Training Company

My internship experience at Bayer AG, located in Monheim am Rhein, Germany, took place within a company recognized for its significant contributions to the life sciences industry. Established in 1863, Bayer has consistently been at the forefront of innovations in healthcare and agriculture, working diligently to improve lives worldwide. Its operations cover key areas such as pharmaceuticals, consumer health, and crop science, establishing its role as a leader in these fields.

Bayer employs approximately 100,000 people globally, each playing a role in the company's wide-reaching impact. This vast workforce is structured under a well-organized hierarchy that fosters a culture of teamwork and innovation. The organization is divided into three main divisions—Pharmaceuticals, Consumer Health, and Crop Science—each with departments focused on advancing Bayer's mission.

The company's diverse product range includes pharmaceuticals with groundbreaking treatments, consumer health products for everyday needs, and crop science innovations for better agricultural outcomes. Bayer serves a varied clientele, from patients and consumers to farmers, all benefiting from the company's commitment to sustainable and impactful solutions.

At the core of Bayer's culture is a dynamic workplace where creativity, collaboration, and diversity are integral. This environment is shaped by the company's fundamental values of innovation, sustainability, and responsibility. Bayer's dedication to employee development through extensive training and advancement opportunities underscores its commitment to fostering a culture of ongoing learning and innovation.

My engagement in the Crop Science division provided deep insights into Bayer's efforts to revolutionize farming through technological and data science applications. Here, the goal is not only to enhance productivity but also to redefine agricultural practices for future sustainability. Initiatives such as developing predictive models and improving crop management practices demonstrated the significant role of data science in advancing agriculture. The division's commitment to integrating technology with farming not only addresses immediate challenges but also lays the groundwork for a future where agriculture continues to support humanity in an environmentally conscious manner.

Looking back on my experience at Bayer, it's clear that the internship was an invaluable period of personal and professional growth. Bayer's comprehensive approach to innovation, coupled with its contributions to society, has made a lasting impact on me. This internship not only expanded my understanding of the possibilities within agricultural data science but also deepened my commitment to contributing to meaningful advancements. As I move forward, I take with me the skills, knowledge, and experiences gained, as well as a deep gratitude for the chance to be part of an organization that genuinely strives to make a difference.

## Job Search and Interview Experience

I discovered the internship opportunity at Bayer AG through LinkedIn, where the company regularly posts openings for various positions. Upon applying, I found the interview process to be smooth and efficient. The individuals involved were highly cooperative, taking the time to understand my internship objectives and ensure a good fit within the organization. Following the interviews, finalizing the contract and setting the start date took an additional 1-2 months, during which Bayer's team remained responsive and supportive throughout the process.

Overall, I am highly satisfied with the journey leading to securing this internship, which was greatly facilitated by Bayer's commitment to engaging and accommodating potential candidates. This reflects the company's dedication to fostering a collaborative and inclusive work environment, further enhancing my enthusiasm for joining the team.

# Progress Report

## 1 Introduction to My Journey

At the start of my internship at Bayer, I had a typical expectation of securing an intern position and working on uncomplicated tasks under close supervision. This perspective was influenced by the popular notion that an intern's role is mostly observational, with minimal potential for significant contributions. Nonetheless, this preconception was quickly nullified as I explored the lively and cooperative

atmosphere of Bayer.

From the outset, my interactions with my supervisor were a highlight, breaking away from the typical hierarchical structure often seen in intern-supervisor relationships. Instead, we engaged in regular, meaningful dialogues that went beyond mere status updates. These check-ins were interactive, focusing on reviewing my progress, addressing challenges, and sharing feedback. Such a collaborative approach fostered a sense of belonging and actively involved me in projects, transforming every task into an opportunity for real contribution and professional development.

The level of responsibility and trust Bayer placed in me was both unexpected and rewarding. Being entrusted with sensitive data and playing a role in key projects, like enhancing predictive models, expanding our databases, and creating unit tests for software applications, highlighted the significant trust and value Bayer sees in its interns. Coupled with a supportive team environment and a shift towards more agile organizational practices, my internship journey became an enriching experience of learning and growth. This experience not only surpassed my initial expectations but also fundamentally changed my perception of the intern role, establishing a new standard for what I seek in my future career paths.

## 2 Adapting to Bayer's Environment

My internship at Bayer began with a burst of activity and adjustment. The intensive onboarding process provided me with a thorough understanding of Bayer's dynamic culture, its ethical framework, and the intricate organization structure, all of which prepared me for my role as a Machine Learning Intern.

This phase of my journey was marked by a series of challenges that tested my flexibility and resilience. From navigating the integration of cutting-edge technologies to working alongside diverse, multidisciplinary team, each day presented a new learning opportunity. The responsibility of managing sensitive data brought to the forefront the critical nature of ethical behavior and professionalism in the workplace.

These experiences underscored the necessity of adaptability within Bayer's fast-paced and evolving environment. Each challenge was an invitation to grow, push-

ing me to not only adapt to new situations but also to thrive within them.

# 3 Core Projects and Contributions

## 3.1 Predictive Modelling for Formulation Stability

Within the Computational Life Science (CLS) Formulation Technology team at Bayer AG, I embarked on a significant initiative aimed at advancing our predictive capabilities regarding the stability of formulations used in agriculture. This initiative was essential for optimizing the performance of herbicides and fungicides critical to global farming operations.

### Enhancing Predictive Analysis in Crop Science

In my role within the CLS-FT team, I was tasked with a specific and intricate challenge: to predict emulsion separation in Emulsifiable Concentrate (EC) formulations. This property is a critical indicator of the overall stability of ECs, which are fundamental to Bayer's product lineup. Addressing this challenge involved dealing with diverse and occasionally inconsistent data accumulated from extensive laboratory experiments over the years. My goal was to distill this complex dataset to accurately forecast this crucial aspect of product stability.

### Project Overview and Objectives

The ambition of the project was to navigate through combined dataset and develop a model that could reliably predict emulsion separation in EC formulations. Ensuring the predictability of this property was vital for upholding the integrity and reliability of Bayer's offerings to the agricultural community.

### Analytical Techniques and Tools Employed

Utilizing Python and the PyTorch framework, I embarked on an exhaustive exploration of the collected data. This journey entailed thorough data analysis processes, including cleaning and visualization, to prepare the groundwork for our predictive model. This effort required a collaborative spirit, engaging closely with

team members and laboratory staff to ensure the accuracy and applicability of our data to the project's rigorous standards.

**Key Insights and Achievements**

This endeavor led to the creation of a predictive model focusing solely on emulsion separation. Although the model didn't achieve the highest levels of accuracy initially hoped for, it provided valuable predictions about this essential stability indicator. My application of a binary classification model helped navigate through the data intricacies, laying a foundation for the model that, despite its modest outcomes, offers optimism for further refinement and improvement.

**Future Directions and Recommendations**

The preliminary results from this initiative shed light on the significant potential for enhancing predictive models in agricultural science, especially concerning EC formulation stability. Future efforts will benefit from broadening the scope of data and possibly integrating additional stability indicators to maintain the model's accuracy. A joint effort to standardize data collection methodologies across laboratory practices will be key to enriching the dataset for more sophisticated modeling work. This pathway of integrating data science with agricultural innovation promises to unfold new solutions in crop protection, thereby enhancing sustainability and productivity in the agricultural sector.

## 3.2 Data Expansion Project

### 3.2.1 Introduction

During my internship, I undertook a project focused on refining and expanding the data repository used by our team. The initiative aimed at enhancing the quality and breadth of data available, primarily through extracting information from existing resources and ensuring its integration into our current database system. This work, though concentrated in scope, was vital for supporting the team's ongoing research and development efforts.

### 3.2.2 Methods Used

Utilizing online sources and extracting data from PDF documents already within our database were the main methods I employed. This process involved a significant amount of manual effort to ensure the newly added data was both consistent and accurate. Given the volume of manual work, I focused on maintaining the highest possible level of precision in my task execution, though the challenge of ensuring absolute accuracy was ever-present.

### 3.2.3 Insights into Operational Efficiency

This project was not just a test of patience and precision in data management; it was a deep dive into the operational backbone of research efforts. I came to understand the critical balance between manual diligence and the need for automated solutions where feasible. The process illuminated the vital role of data integrity in research and development, underscoring that even the most granular task, when executed with care, can significantly enhance a project's foundation.

### 3.2.4 Carrying Forward Principles of Data Integrity

My contributions, though initially appearing modest, held significant value in refining our data repository and facilitating more efficient access to high-quality data. This groundwork has the potential to greatly accelerate future research initiatives within the team. This experience made it clear how important careful data management is to getting the results you want from your research. Moreover, it instilled in me a profound sense of responsibility towards maintaining the integrity of research data, a principle I will uphold throughout my career.

## 3.3 Software Quality Assurance Through Unit Testing

### 3.3.1 Testing Frameworks and Strategy

My final project during the internship involved a deep dive into unit testing, focusing on the software packages I had been working on. This task was essential for ensuring the quality of our software. Using Python's testing frameworks, unittest and pytest, I began by thoroughly understanding the code I was responsible for.

This initial step was critical, as it required grasping how the code operates, its intended outcomes, and potential areas where errors could emerge. This kind of preparation was essential to forming a new way of thinking that was focused on comprehensive testing.

### 3.3.2 Notable Findings and Fixes

The process of unit testing was an enlightening experience. It highlighted the importance of being intimately familiar with the code's expected behavior. At times, the task of identifying and rectifying errors was challenging, but it became more manageable with guidance from my supervisor. Through this detailed examination, I discovered several bugs that, if left unchecked, could have led to significant issues down the line. Correcting these problems not only improved the current functionality but also made the software more robust against future errors. This part of my internship also improved my proficiency with Git and GitHub, which were invaluable for managing the changes I made.

### 3.3.3 Reflecting on the Process

Delving into unit testing for my final internship project was a comprehensive lesson in the importance of quality assurance in software development. Learning to anticipate potential problems and understanding the deeper workings of the code was a vital part of this project. The ability to consult with my supervisor to confirm the expected behavior of the code was crucial. Addressing the identified bugs helped improve the overall quality of the software packages our small team was developing.

This task underscored the significance of preparing and adopting a focused approach to unit testing. It wasn't just about finding errors; it was about ensuring the software performed as expected in every scenario. This project, being my last during the internship, marked a significant culmination of my learning experience, highlighting the growth I've achieved in understanding and applying software quality assurance practices.

## 3.4 Knowledge and Skills Acquired

My internship journey was a truly engaging experience that expanded my knowledge and skills in many areas. I got to develop in ways that rounded out my professional abilities, making me better prepared for my future career.

### 3.4.1 Improving Data Science Skills

During my time, I worked on enhancing models and diving deep into data science, which significantly boosted my understanding and skills. Even though I didn't build a PyTorch model from scratch, working on extending one helped me understand its workings in depth. This was crucial for my growth in handling complex data effectively. Also, getting familiar with Git and GitHub played a big part in improving my ability to manage projects and collaborate with others. Microsoft Office was another tool that became essential, especially Excel, for organizing and analyzing data.

### 3.4.2 Learning About Database Management

My work involved collecting and organizing data, where I gained practical experience with database management. While I mainly worked with other tools and not MariaDb, the principles I learned are broadly applicable, helping me appreciate the importance of accurate and well-organized data for research and insights.

### 3.4.3 Gaining Insights from Daily Tech Talks

Although I wasn't an active participant in daily tech talks, listening in was incredibly beneficial. It was a great way to absorb new information and learn from others' experiences and expertise. These sessions were an excellent resource for broadening my knowledge base beyond my direct tasks.

### 3.4.4 Enhancing Soft Skills through Team Interaction

The team meetings and daily lunches were invaluable for improving my soft skills. These interactions provided a clear view of how agile methodologies are applied in real workplace settings, enhancing my understanding of effective teamwork and

communication. Being part of a team that focuses on data insights and research was rewarding, and it was fulfilling to support the work and contribute to our shared goals.

### 3.4.5 Cultivating Time Management and Responsibility

The ability to manage my time well and have a strong sense of accountability for my work were essential to my professional development. I learned how to prioritize and effectively manage my time by navigating the difficulties of meeting project deadlines and juggling several tasks at once. This skill was further honed through the autonomy I was granted, pushing me to take ownership of my projects and ensuring their successful completion.

## 4 Conclusion

In summary, my internship at Bayer has been a journey marked by significant growth and achievement. I adapted seamlessly to Bayer's dynamic ecosystem, navigating challenges with resilience and flexibility. Through core projects like enhancing predictive analysis, expanding our data repository, and delving into unit testing, I've not only honed my technical skills but also gained practical insights into the intricacies of data management and software quality assurance.

Furthermore, interactions with my supervisor and team members, coupled with daily tech talks, enriched my understanding of effective teamwork and agile methodologies.

Overall, this internship has provided me with a solid foundation for my future career, equipping me with the tools and experiences necessary to thrive in the professional landscape.

## Final Assessment

During my internship at Bayer AG, I learned a lot, both professionally and personally. This wasn't just about experiencing Bayer's innovative culture but also about overcoming real challenges, which helped turn my academic knowledge into practical skills.

I worked on several projects that expanded my understanding and abilities, especially in data analysis, which applies to more than just agriculture. These tasks pushed me to be creative and improve my technical skills. Seeing the results of my work was incredibly rewarding and fueled my interest in data science.

Bayer's supportive environment was crucial for my growth. It allowed me to learn new technologies, improve my skills, and understand the value of teamwork. Working with diverse teams taught me about cooperation and the importance of different perspectives.

This internship was a journey of self-discovery for me. I've become more adaptable, confident, and ready to face new challenges. Bayer's culture of aiming for excellence inspired me to continuously seek improvement.

Reflecting on my experience, Bayer has shown a strong commitment to making a difference, which aligns with my values. As I move on, the skills and knowledge I've gained are not just for agriculture but can be applied broadly. I'm grateful for the chance to work on meaningful projects and be part of a team that values innovation and collaboration. This experience has prepared me well for future opportunities and has instilled in me a desire to make a positive impact wherever I go.