DELIVERABLE 1.8 - MACHINE LEARNING OF BIOLOGICAL DATA

JULY 28th 2025

biont-training.eu

contact@biont-training.eu

Co-funded by
the European Union

About this report

Grant Agreement	101100604 - DIGITAL-2022-TRAINING-02
Action Acronym	BioNT
Action Title	Bio Network for Training
Deliverable	1.8 Machine learning of biological data
Work package	1 Training design and development
Dissemination level	Public
Authors	S. Di Giorgio, T. Müller, I. Paredes Cisneros, S. Razick, P. Samarakoon, approved by all BioNT partners
Delivery date	2025-07-28

Consortium members

Acronym	**Partner**
EMBL	EUROPEAN MOLECULAR BIOLOGY LABORATORY
BIOBYTE	BIOBYTE SOLUTIONS GMBH
HPCNOW	HPC NOW CONSULTING SL
UO	UNIVERSITETET I OSLO
UB	UNIVERSITAT DE BARCELONA
ZBMED	INFORMATION CENTRE FOR LIFE SCIENCE
RIcapacity	RIcapacity
ALU-FR	ALBERT-LUDWIGS-UNIVERSITAET FREIBURG
EPFL	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE

Table of Contents

Project Overview

The BioNT consortium is dedicated to providing a comprehensive training program and fostering a community for digital skills relevant to the biotechnology industry and biomedical sector. With a curriculum tailored for both beginners and advanced professionals, BioNT aims to equip individuals with the necessary expertise in handling, processing, and visualising biological data, as well as utilising computational biology tools. Leveraging the consortium's strong background in digital literacy training and extensive network of collaborations, BioNT is poised to professionalise life sciences data management, processing, and analysis skills.

This Workshop at a glance

The seventh BioNT training workshop, Applied Machine Learning for Biological Data, comprised a pre-workshop industry meetup event followed by the main workshop, which was divided into Module 1 and Module 2 (detailed later in the report).

The industry meetup event, titled "AI in Biomedical Data - Industry Seminar: Bridging Innovation & Opportunity", aimed to engage our target audience and raise awareness of the BioNT project, particularly the "Applied Machine Learning for Biological Data" workshop. The event was designed for researchers, healthcare professionals, industry leaders, and policymakers interested in the intersection of Artificial Intelligence (AI) and biomedical sciences. The meetup highlighted the "Applied Machine Learning for Biological Data" as a valuable opportunity to gain introductory knowledge and skills relevant to those transitioning into the field of AI in biomedical data. It also provided a platform for participants to discuss key topics and offer feedback, which was subsequently used to adapt Module 2 of the workshop. Approximately 75 participants attended this event either in person or via a streaming service. Additionally, the LinkedIn event we published afterwards received over a thousand views.

Key discussion topics of the industry meetup event:

1. AI-driven innovations in medical research, diagnostics, and personalised medicine.
2. Data integration and interoperability challenges within healthcare systems.
3. Ethical and regulatory considerations for AI in biomedical applications.
4. Industry-academia collaboration to accelerate the adoption of AI in healthcare.
5. BioNT's training offerings for competence building in biomedical data science and AI.

The "Applied Machine Learning for Biological Data" workshop was organised into two modules: an optional Module 1 and a mandatory Module 2. Module 1 was scheduled as two half-day sessions (May 27th and 28th, 2025) and focused on essential data handling techniques using NumPy and Pandas, which are tools widely adopted for manipulating and analysing biological data. This optional module was open to anyone interested, not just the registered participants of the main workshop. It aimed to bridge the knowledge gap between the Python course provided in BioNT’s basic curriculum and the Python skills required to follow the advanced Machine Learning (ML) topics covered in Module 2.

Module 2 was scheduled for 5 full days (June 2nd to 6th, 2025) and covered a broad range of ML topics. These included core ML concepts with short exercises, extended hands-on sessions applying ML to biological use cases, as well as theoretical knowledge and practical sessions on GPU-powered genomics workflows.

Feedback from participants indicated exceptionally high satisfaction levels across all measured dimensions (as detailed in the section “Post-workshop Survey”). The workshop's overall success was underscored by recommendation likelihood scores: 92% participants rated it between 8-10, with 67% participants giving the maximum score of 10, indicating they would very likely recommend the workshop to colleagues.

As further detailed in later sections of this report, the workshop successfully delivered high-quality instruction in a supportive learning environment with strong practical relevance for participants.

Training materials, resources and advertising

As this workshop was delivered for the first time, both the lecture materials and hands-on sessions were developed from scratch. The lecture content was specifically tailored to address biological datasets and examples, while the hands-on sessions focused on research and clinical use cases in Genomics. These materials were designed to align with BioNT's mission of bridging academic and industrial needs and practices. All materials are openly available and licensed under the MIT License, enabling future reuse and adaptation by other training communities.

Day and session	**Topics**
Module 1	
Day 1: NumPy: Fundamentals	Introduction to NumPy NumPy Data Types Array Indexing and Slicing NumPy Boolean Masking and Filtering Essential Array Operations with NumPy Vectorised Operations in NumPy: Beyond Python Loops Hands-on homework: RNA Expression Analysis - alternative method
Day 2: Pandas: Fundamentals	Introduction to Pandas Data Import and Export in Pandas DataFrame Manipulation & Sorting Indexing, Selection & Slicing in Pandas

	Summary Statistics & Aggregations in Pandas Hands-on: RNA Expression Analysis - alternative method
Module 2	
Day 1; Morning session	Introduction to Module 2 ML terminology and ML in Bioinformatics Unsupervised Learning: Clustering (K-Means Clustering, Hierarchical clustering, Clustering evaluation metrics)
Day 1; Afternoon session	Unsupervised Learning: Dimensionality reduction (Principal component analysis - PCA) Hands-on session demonstrating PCA and clustering in cancer genomics
Day 2; Morning and afternoon sessions	Classification: Logistic regression; Tree-based methods; Matrices for classification evaluation Hands-on session demonstrating Logistic regression in cancer genomics
	Regression: Regression mechanics, Loss function, Regularised regression, Matrices for regression evaluation
Day 3; Morning session	Model validation and optimisation (Overfitting and underfitting, Standardising Data, Handling missing data)
Day 3; Afternoon session	Model validation and optimisation (K-fold cross-validation) Hands-on session: ML workflow with biological data
Day 4; Morning session	Introduction to deep learning (Basic concepts of Neural Networks - NN; Simple NN with PyTorch)
Day 4; Afternoon session	Building simple NN with PyTorch, Deep learning applications in genomics Hands-on session demonstrating deep-learning-based variant calling via DeepVariant
Day 5; Morning session	Introduction to Accelerated Genomics (NGS data analysis, GPU introduction) GPU introduction Docker introduction
Day 5; Afternoon session	Hands-on session on implementing accelerated Genomics workflows with Parabricks on VM with GPUs

Advertisement

The workshop was advertised via social media, several websites, mailing lists and Slack spaces of networks and communities (ELIXIR, BioRN cluster, NFDI4Microbiota and de.NBI, among others). For the advertising of this workshop, a tailored image was generated, which included a QR code to facilitate access to the registration platform, as shown in Figure 1.

A.

BioNT
887 followers
1mo •

Are you interested in learning how to use NumPy and Pandas to handle biological data? Or you're planning to attend Module 1 of our Machine Learning course?

We've got you covered!
Here are the instructions to set up your computational systems and the zoom link so you can fully participate in the course.
And the best part? No registration is required!

https://lnkd.in/eEvbpim9

#bioinformatics #Machinelearning #DigitalLifeSciences
#BioinformaticsWorkshop

B.

BioNT
887 followers
2mo • Edited •

🚀 Are you a bioinformatician working in a #StartUp or #SME 🚀, and eager to dive into the world of Machine Learning? 🤖
If so, we have exciting news for you!

Join BioNT's free online workshop on Applied Machine Learning for Biomedical Data, happening at the end of May/early June 2025.

By the end of the workshop, you will be able to:
💡 Communicate and collaborate more effectively with Machine Learning experts.
💡 Understand core concepts of classification and regression.
💡 Build, train, and evaluate machine learning models and simple neural networks.
💡 Implement GPU-accelerated pipelines to scale your bioinformatics workflows.

🔗 Discover the full program and apply now — Seats are FREE but limited!
https://lnkd.in/eiqTesB8

⚙️👥 Workshop Format: The sessions will run as Zoom webinars — you'll be able to learn directly from expert trainers, and if you prefer, you can stay completely anonymous throughout the workshop.
Don't miss this opportunity to upgrade your skills and bring new capabilities to your SME!
#bioinformatics #Machinelearning #StartUp #SME #training #AI
#digitaltransformation #careergrowth

Figure 1 - Module 1 (A) and Module 2 (B) of the workshop advertised through LinkedIn.

Technical infrastructure

Module 1 ran over two morning sessions, from 09:00 to 12:00 CEST on the 27th and 28th of May 2025 and did not require prior registration. Module 2 was scheduled over 5 days with sessions delivered from 09:00 to 16:00 CEST from June 2nd to June 6th, 2025, and required registration through the event website. The entire event (Modules 1 and 2) was conducted virtually and offered at no cost to participants.

Workshop webpage, participant registration and selection

The CECAM event management platform, provided by the EPFL, was used to create a dedicated webpage for the workshop. This webpage included the workshop description, learning objectives, requirements, program, and other relevant information for potential participants. This event webpage with the necessary information is accessible at: https://www.cecam.org/workshop-details/applied-machine-learning-for-biological-data-1459, where the full program, learning outcomes and requirements are listed.

Registration was required only for Module 2 of the workshop (discussed later in this section), and the CECAM platform was used to manage the applicant's information and communication. In parallel, the EMBL servers were used to collect pre- and post-workshop information through pseudo-anonymised surveys. Survey responses were linked to applicant data only via a unique identifier, which was provided during the CECAM registration process and entered again in the EMBL-based survey. This approach ensured that only the workshop organisers had access to identifiable applicant data while still collecting information relevant to the workshop separately. To register, applicants had to: (i) register on the CECAM platform, (ii) complete and submit the pre-workshop survey, and finally (iii) complete the application on the CECAM platform using the unique identifier provided in the pre-workshop survey.

Module 2 of the course, which focused on hands-on ML implementation and accelerated genomics workflow implementation, required participants to use virtual machines equipped with Graphical Processing Units (GPUs). Due to high costs, course organisers could only provide 40 VMs, necessitating a selection process from the 125 applicants for Module 2. Module 1 remained open to all interested participants as it did not require VM resources. The selection process for Module 2 was based on the answers provided in the pre-workshop survey (containing no personal information). For the selection criteria for Module 2, learners currently working in European countries who were either job seekers, small-and-medium size enterprise (SME) employees, SME collaborators, or individuals aiming to work in SMEs were prioritised. Among applicants meeting these criteria, participants were selected on a first-come-first-served basis to ensure fair and transparent allocation of the limited computational resources.

The 40 selected participants for Module 2 represented mostly European countries, as given by the selection criteria explained above. Career-wise, the cohort was dominated by experienced and early-career academics, with postdoctoral researchers and graduate students forming the majority, alongside research staff and analysts. Regarding SME engagement, an overwhelming number of applicants expressed aspirations to work in SMEs (SME job seekers), while the rest either collaborated with SMEs or were current SME employees. Selected participants showed strong alignment with biological data applications, particularly in genetics, genomics, bioinformatics, life sciences, and biomedical/health sciences, followed by high-performance computing expertise. This composition highlighted the selection process's effective targeting of the desired demographic, aligning with the project's goals.

Infrastructure for the workshop

Zoom

The workshop was delivered via Zoom webinar, allowing participants to learn directly from the trainers. To facilitate video recording for future self-paced learning materials, participants' microphones were muted, and cameras were off during presentations. All direct interactions were instead conducted in written form via a prestructured collaborative document, which allowed for anonymous participation. To serve the collaborative documents, a HedgeDoc collaborative space was set up by BIOBYTE, and was hosted on their server.

To further enhance the interactive experience, the instructors occasionally engaged in dialogue around the topics under discussion. The idea was to bring in some of the “real world” experience of the instructors when it came to the topics, and to show that the simplified examples of the hands-on can have a practical impact in larger projects.

Shared documents

A **Main** collaborative document, set up by the instructors and organisers, was shared with the participants before the workshop. Each section of the workshop had dedicated *Hands-on* boxes to report on the task status, ask questions or raise issues. Helpers engaged and assisted participants by answering the questions and issues directly in this document. This Main document was updated live during the workshop. Separate boxes to answer questions were used to improve participant engagement and as an indirect learning assessment (Figure 2).

✏️ **Hands-on: Access this HedgeDoc main document**

❓ **Are you on this HedgeDoc? (Add a + when done)**

- Yes:
- No (please sent a e-mail to one of the helpers)

❓ **Do you need help? Please describe your issue**

-
-
-
-

❓ **Have you ever used Markdown? (Add a +)**

- Yes
- No
- What is Markdown?

❓ **We plan to adjust the break times according to the workshop's progression. Does this arrangement work well for everyone?**

- Yes
- No

Figure 2 - Example of a question and hands-on boxes in the collaborative document.

To help with the organisation, four HedgeDoc documents were used: (i) a **Templates** (Module 1 and Module 2) with all instructions and boxes for hands-on, questions, etc; (ii) the **Main** documents (Module 1 and Module 2) with the information for the participants during the workshop, filled with boxes related to the section covered by the instructor to help with the navigation and cleaned during each break to avoid an overcrowded document; and (iii) a **History** documents (Module 1 and Module 2) collecting all the content from the Main document. This document was shared with participants during the workshop to grant them access to all prior conversations. In addition, (iv) a document for Helpers & Instructors (Module 1 and Module 2) was created, providing the workshop setup, interactions and explaining tasks of the helpers and instructors.

There was a very lively interaction with the learners through the shared document. This was made possible by the extended set of helpers available to answer participant questions.

Instructors, from time to time, used the shared document to check the satisfaction with lesson delivery and made necessary adjustments. Type-along sessions, where the learners tried to follow the instructor on their computers, and demo sessions, where the learners watched the instructor completing tasks, were clearly communicated to the learners.

Technical software tools, programming modules and specialised hardware

The workshop required a significant set of development tools to be available to the users. These were defined as a set of prerequisite software and tools (Module 1 and Module 2 installation guides), which included Bash, Python and Jupyter Notebooks (with both Linux shell and Python experience being course prerequisites). It also required them to set up their Python environment according to precise instructions provided in the installation guides.

Trainees: Demographics, expertise and expectations

Following the selection process, we extracted the pre-workshop survey responses provided by the selected 40 participants. These responses were then analysed and shown in the following sections. The pre-workshop survey comprised 22 questions covering skills, demographics, and miscellaneous topics. The majority of questions were optional for the successful completion and submission of the survey.

General information

Of the 40 selected participants, 19 were male, 18 female, 1 identified as non-confirming and 2 preferred not to specify. Nationality and employment are summarised in Figure 3.

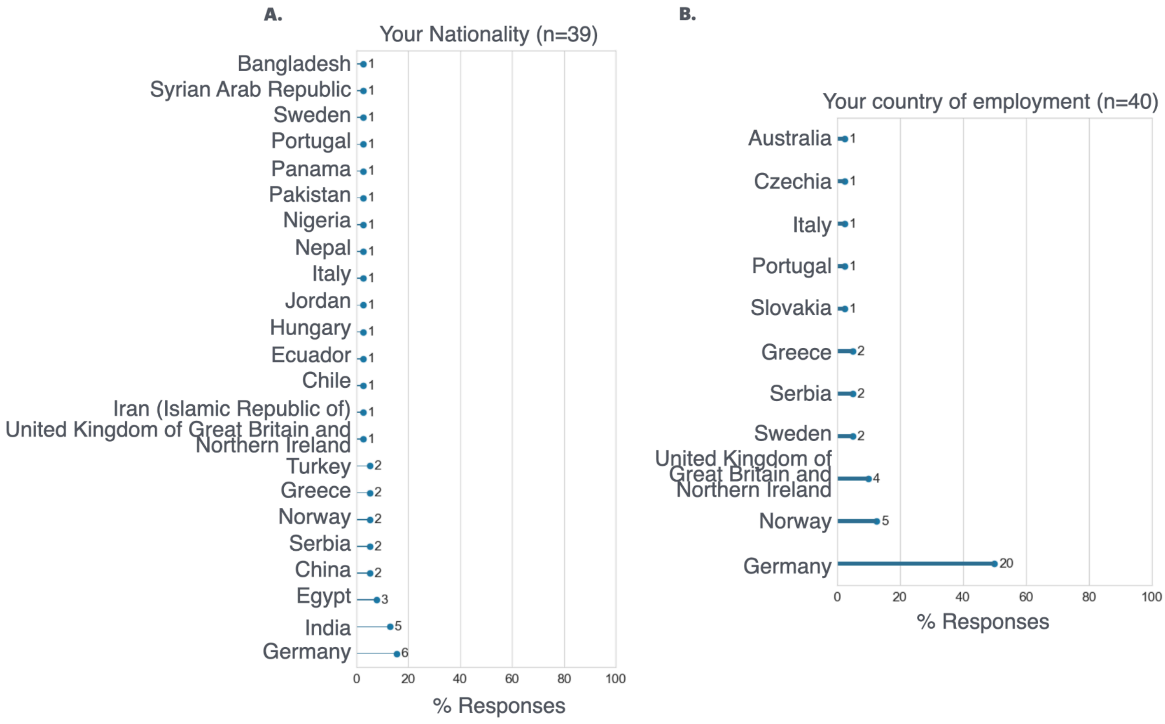


Figure 3 - Workshop applicants' nationality (A) and country of employment (B) from the pre-workshop survey.

As previously described in this report, participants were pre-selected primarily based on their country of employment, prioritising those based in Europe, with one individual from Australia included to honour a specific collaborative commitment from BioNT. Most of the selected participants were from Germany, followed by notable groups from Norway and the United Kingdom.

During the planning of the AI Industry Meetup, a misunderstanding occurred with one invited speaker, who believed they had confirmed their participation as a presenter. However, no formal confirmation had been received by the organising team, and no follow-up communication took place. The speaker later acknowledged that the confusion may have stemmed from previous correspondence with a different organisation and expressed understanding of the situation. In response, and as a gesture of goodwill, BioNT offered a reserved seat in its upcoming online ML course to a member of the speaker's affiliated organisation. This seat was ultimately taken by a PhD student based in Australia. This incident reflects BioNT's commitment to maintaining positive relationships with its partners and supporting the professional development of early-career researchers, even in cases of logistical miscommunication.

Interestingly, while employment was largely Europe-focused, participant nationalities were globally diverse. Germany also stood out as the largest single nationality group, followed by a substantial number of participants from India. Other participants came from countries across Europe, Asia, and Africa, reflecting a rich international mix within a predominantly European employment context. This distribution highlights the targeted geographic selection while still embracing a broad spectrum of global talent.

The survey participants represent a highly specialised group, primarily concentrated in the field of life sciences and computational biology (Figure 4). Nearly three-quarters of respondents (~76%) reported working in genetics, genomics, or bioinformatics, with strong additional representation in the broader life sciences (~58%) and biomedical or health sciences (~55%). High-performance computing was also well represented (~26%), underscoring the computational demands of contemporary biological research.

Notably, there was little to no representation from the humanities, social sciences, or most traditional physical sciences. This indicates that the workshop was highly targeted toward a specific community engaged in data-intensive biological research, rather than a broader cross-section of academic disciplines.

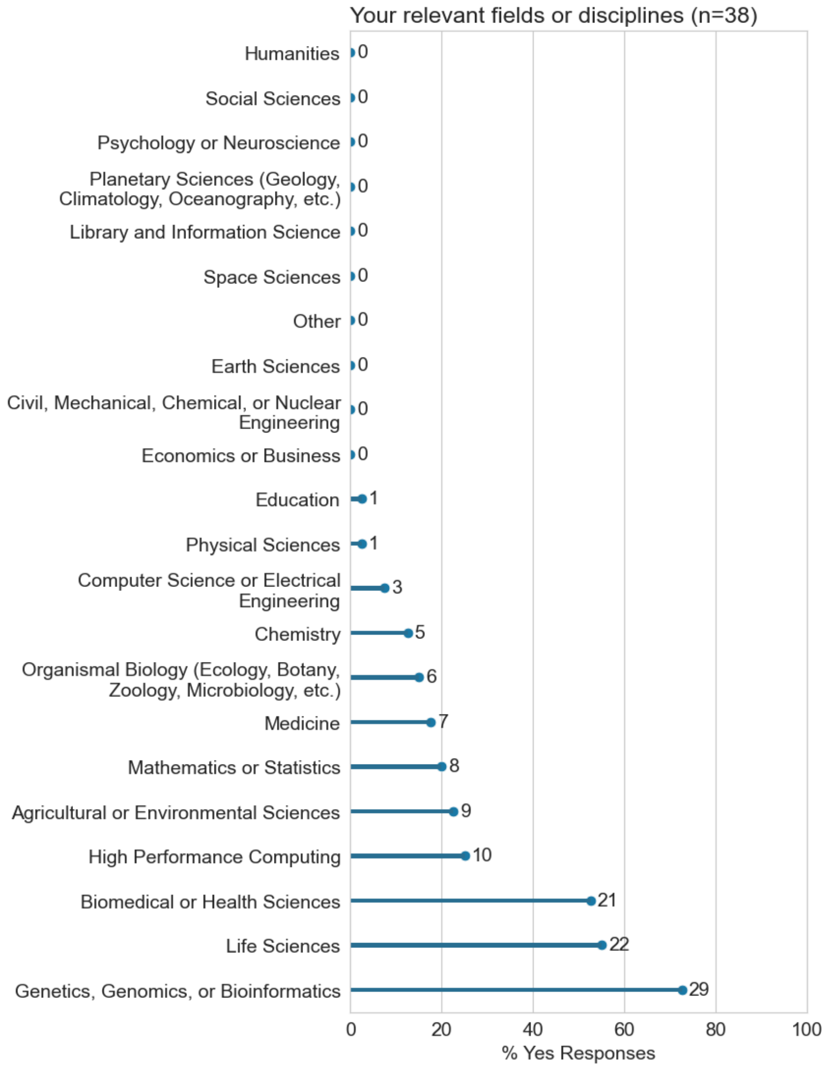


Figure 4 - The relevant fields or disciplines (multiple choice) question pre-workshop survey (n = 38).

The survey participants were predominantly experienced and early-career researchers (Figure 5A), with postdoctoral researchers (~35%) and graduate students (~30%) comprising nearly two-thirds of respondents. Research-focused positions dominate the rest of the group, including research staff and analysts, while only two participants hold faculty roles, indicating limited representation from senior academics. Despite their current employment, participants expressed a strong interest in transitioning to industry careers, particularly within SMEs (Figure 5B). Nearly all participants reported some connection to an SME. Specifically, 3 participants currently work in SMEs, 7 already collaborate with one, and a significant number (n = 29) indicated a desire to work in an SME. This contrast between current academic employment and SME connection suggests that the survey captured researchers collaborating with SMEs or actively seeking to transition from academia to industry.

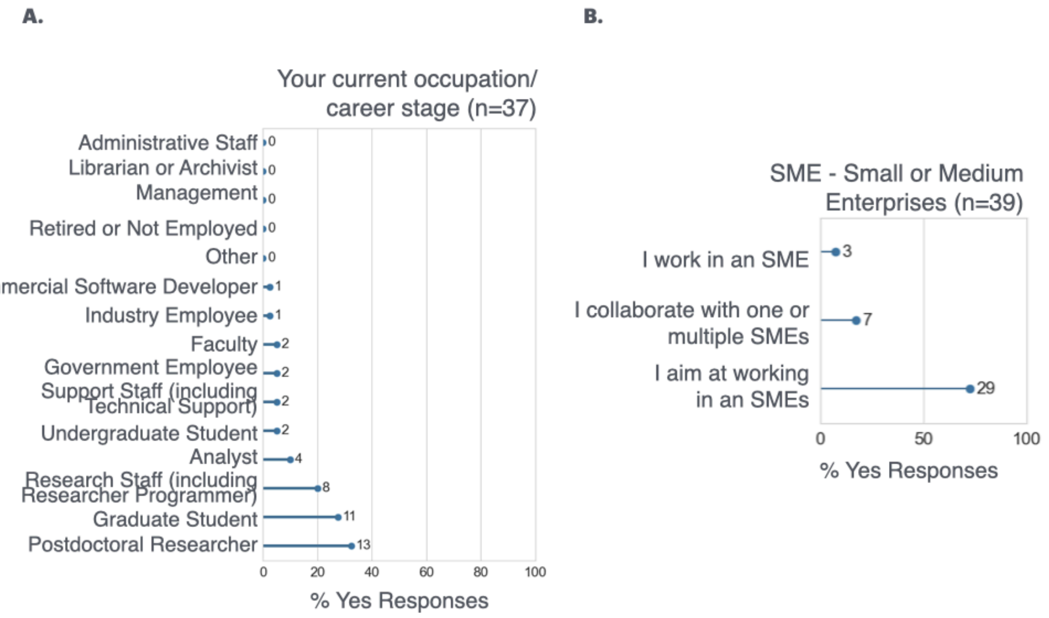


Figure 5 - The current occupation/career stage (multiple choice; n = 37 respondents) (**A**); Connection of the workshop participants to SMEs (n = 39 respondents) (B) according to the pre-workshop survey.

Applicants found information about the workshop through various channels, as illustrated in Figure 6, with the majority learning about it via social media or receiving direct recommendations from friends or colleagues.

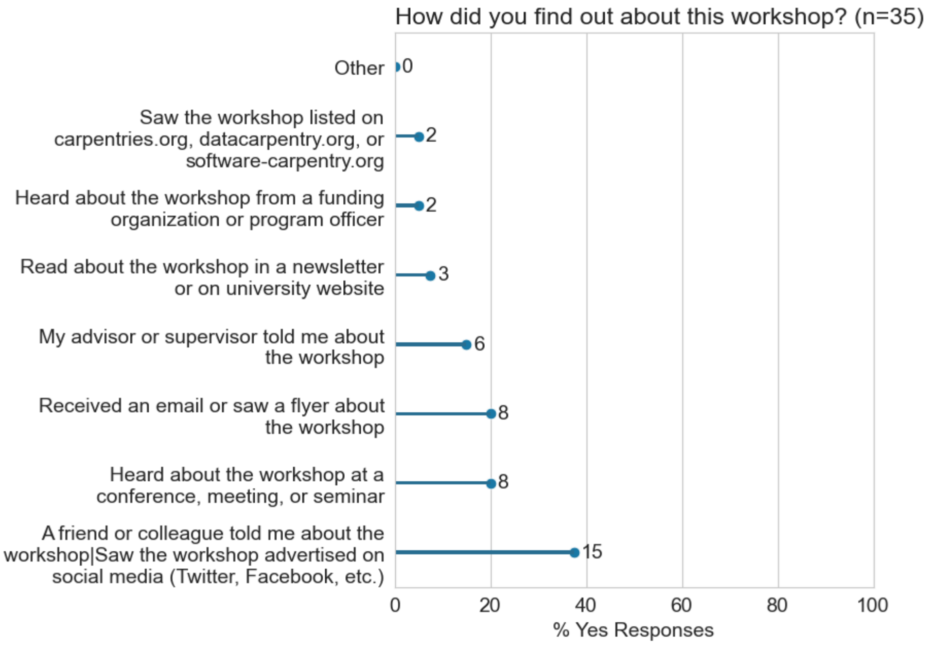


Figure 6 - Answers to the question: “How did you find out about this workshop? (multiple choice)” (n = 35).

The survey data reveals a distinct divide between participants' comfort with command-line and programming tools versus traditional GUI-based specialised software. A majority of participants demonstrate advanced computational skills, with ~61% using programming languages (R, Python, C++, etc.) daily and an additional ~23% using them weekly (Figure 7A). This high percentage of programming language usage (total of daily and weekly usage) aligns with the ~77% who use command shell interfaces daily or weekly, indicating a participant base that is highly comfortable with text-based, technical computing environments (Figure 7B). The strong adoption of version control systems further reinforces this technical specialisation, with 49% using Git or similar tools at least weekly, suggesting collaborative coding practices and systematic project management approaches (Figure 7C).

In contrast, the usage patterns for GUI-based specialised software and database systems show much more varied adoption levels. Only ~13% of participants use specialised software like SPSS, SAS, or GIS tools on a weekly or daily basis, with the majority (~71%) using such tools monthly or less frequently (Figure 7D). Database usage presents a middle ground, with ~38% using SQL or similar systems monthly or more frequently, while ~18% never use databases at all (Figure 7E). This pattern suggests that while participants are highly proficient in programming and command-line environments essential for ML work, they may have less reliance on traditional point-and-click analytical software, potentially indicating a preference for programmatic approaches to data analysis that offer greater flexibility and reproducibility, which aligns with the course’s focus.

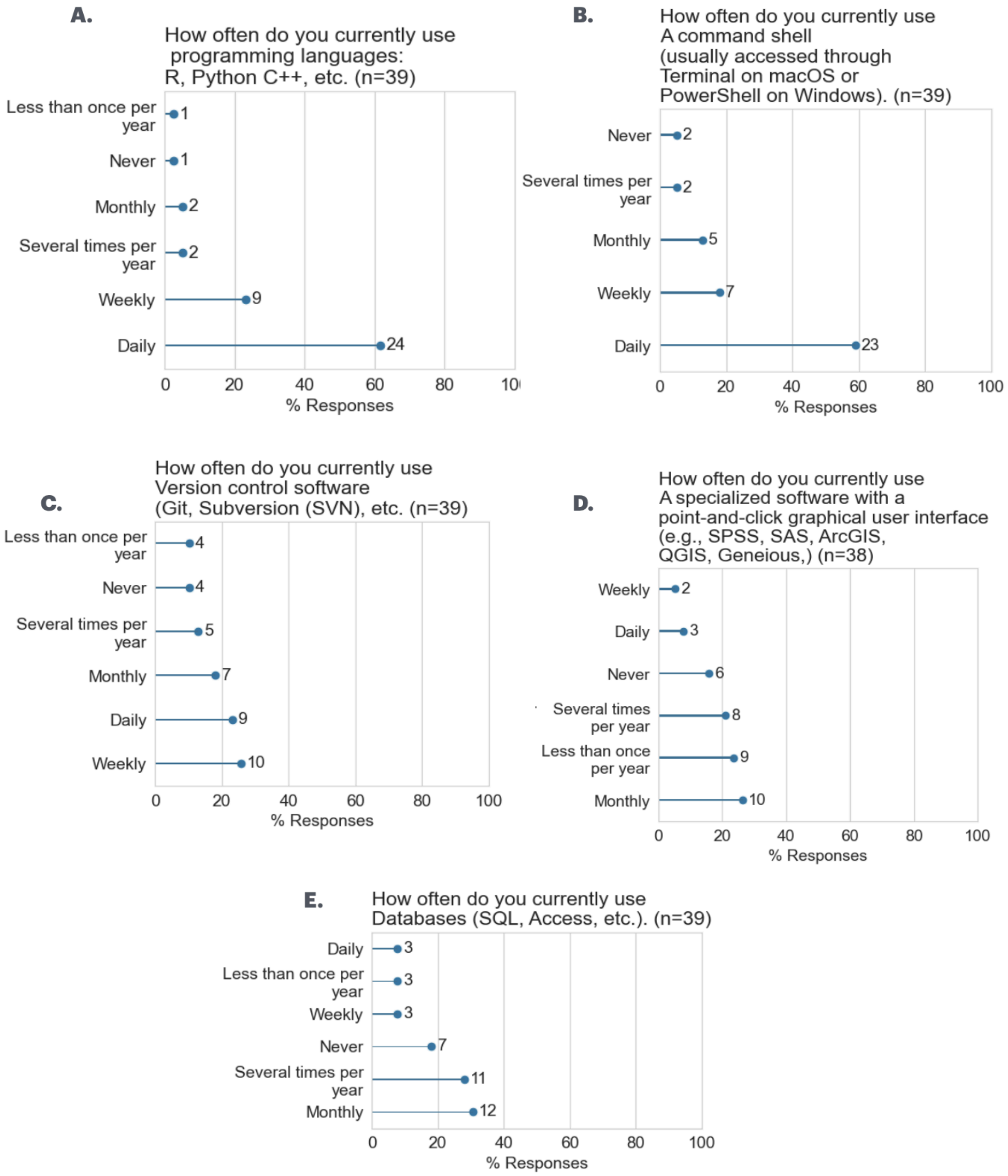


Figure 7 - Answers to the question: “How often do you use any of the following?”. This question was optional and trainees could answer none, some, or all the questions (n = 38 - 39): **A** - Programming languages: R, Python, C++, etc.; **B** - Terminal and macOS or PowerShell on Windows; **C** - Version control software (Git, Subversion (SVN), etc.); **D** - A specialised software with a point-and-click graphical user interface (e.g. SPSS, SAS, ArcGIS, QGIS, Geneious); **E** - Databases (SQL, Access, etc.).

Most participants were keen on acquiring new skills (~80%), with some specifically interested in learning those applicable to their future job (~65%) and current occupation

(~50%). Interestingly, about 55% of the participants expressed their intentions to leverage the acquired skills either to secure a promotion within their current job or to pursue new employment opportunities (n = 22) (Figure 8).

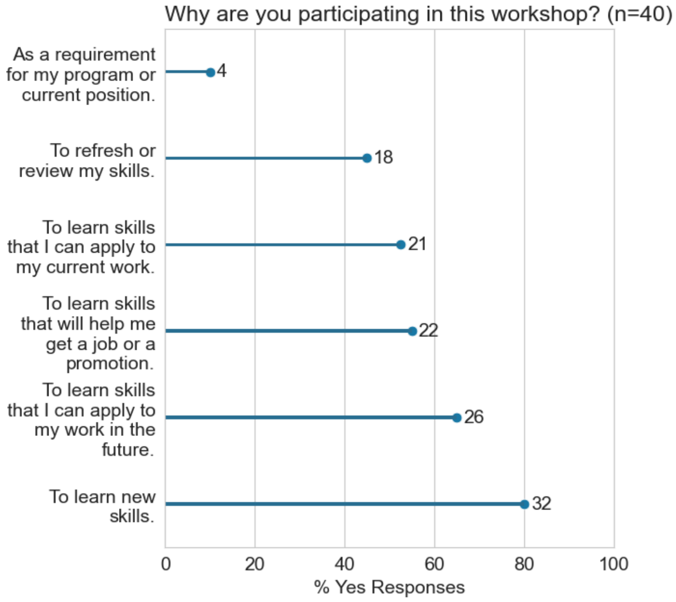


Figure 8 - Answers to the question: "Why are you participating in this workshop? (multiple choice)" (n = 40).

Participation

All 40 selected applicants of Module 2 attended the workshop live, although not all of them stayed for the entire duration (Table 3). All the selected participants received the self-learning materials for them to consult at any time.

Day	Participants	Instructors	Helpers
1	39	3	5
2	34	3	5
3	35	3	5
4	35	2	4
5	30	2	3

Table 3 - Number of participants, instructors and helpers per day. The number of participants was obtained from the participant login information captured by Zoom.

Evaluation

Twenty-six workshop participants submitted the homework assignment and requested a certificate. Submissions were marked against a set of assessment criteria, with a pass mark of 50. All participants who scored above 50 were eligible for a certificate.

Assessment criteria

- Data Preparation: 25 points
    - Optional steps: 10 points (at least two of the following are completed)
        - Handle missing values
        - Perform exploratory data analysis (EDA)
    - Mandatory steps: 15 points
        - Split data into training and testing sets
        - Apply appropriate feature scaling/standardisation
- Model Implementation: 30 points
    - Selection of a model
    - Model training
- Model evaluation: 25 points
    - Cross-validation (5-fold)
    - Final model evaluation on test set
- Optional 20 points
    - Clear variable names
    - Proper comments
    - markdown explanations

In summary, all 26 participants who submitted the assignments scored over 50. Out of 26, two participants scored between 55-65, while the others scored over 75. Therefore, all 26 participants who scored over 50 were issued a certificate (Figure 9).

Course co-organisers

Norwegian AI Cloud

cecam
Centre Européen de Calcul Atomique et Moléculaire

Co-funded by the European Union

BioNT
TRAINING AND PROFESSIONALISATION
IN COMPUTATIONAL LIFE SCIENCES

certifies that

<name>

has successfully completed the course

Applied Machine Learning for Biological Data

held from May 27th-28th and June 2nd-6th, 2025

Sabry Razick Ph.D
BioNT trainer,
Manager Norwegian AI Cloud (NAIC)

<Data>
Date

Figure 9 - Certificate provided to attendees who successfully completed the assignment (scored over 50).

Impact and outcomes

Daily feedback

At the end of each day, participants were asked for feedback on the following three points:

- Please share one thing that was good about today
- Please share one thing that could be improved about today
- Do you have any other comments?

The workshop garnered positive feedback, especially for its well-defined structure, organisation, and content. More specifically, the initial day's hands-on sessions of Module 2 received mixed reviews; some participants valued the in-depth code explanations, while others desired more self-coding time for concept internalisation. Consequently, subsequent hands-on sessions were adjusted to include more self-coding time and quizzes to facilitate the understanding of practical use cases. The feedback from the last few days of the workshop indicated participants' appreciation for changes made in the hands-on delivery.

Post-workshop survey

At the end of the workshop, participants were invited to complete a post-workshop survey consisting of 15 questions, some of which were optional. A total of 24 participants completed the survey.

The evaluation data reveal high satisfaction levels across all measured dimensions, indicating a successful educational experience. For example, 23 out of 24 (95.9%) participants responded with "Strongly Agree" or "Agree" that they felt comfortable with the workshop environment (Figure 10A), creating the foundation for effective learning. This positive atmosphere translated into strong perceived practical value, with 75% of participants confident they can immediately apply the learned material to their work or research (i.e., 18 out of 24 responding with “Agree” and “Strongly Agree”; Figure 10B). The immediate applicability rating is particularly significant given the technical nature of ML concepts, suggesting that the course successfully bridged theoretical knowledge with practical implementation skills that participants can deploy in their biological data analysis workflows.

Instructor effectiveness emerged as a standout strength of the program, receiving consistently positive ratings across all evaluated criteria. For example, survey questions assessing instructors’ clarity of answers (Figure 10C), enthusiasm (Figure 10D), and approachability (Figure 10E) showed that ~96% of participants (23 out of 24) responded positively with “Strongly Agree” or “Agree” answers. The survey also evaluated the instructors' knowledge with the question “The instructors were knowledgeable about the materials being taught”. Again, 23 out of 24 of the participants (~96%) selected “Strongly Agree” or “Agree”, reflecting a high level of confidence in the instructor's expertise (Figure 10F).

Notably, one participant consistently provided negative feedback across all categories, suggesting an outlier experience, while the remaining responses demonstrated consistency in positive evaluation. These results indicate that the workshop successfully delivered high-quality instruction in a supportive learning environment with strong practical relevance for participants.

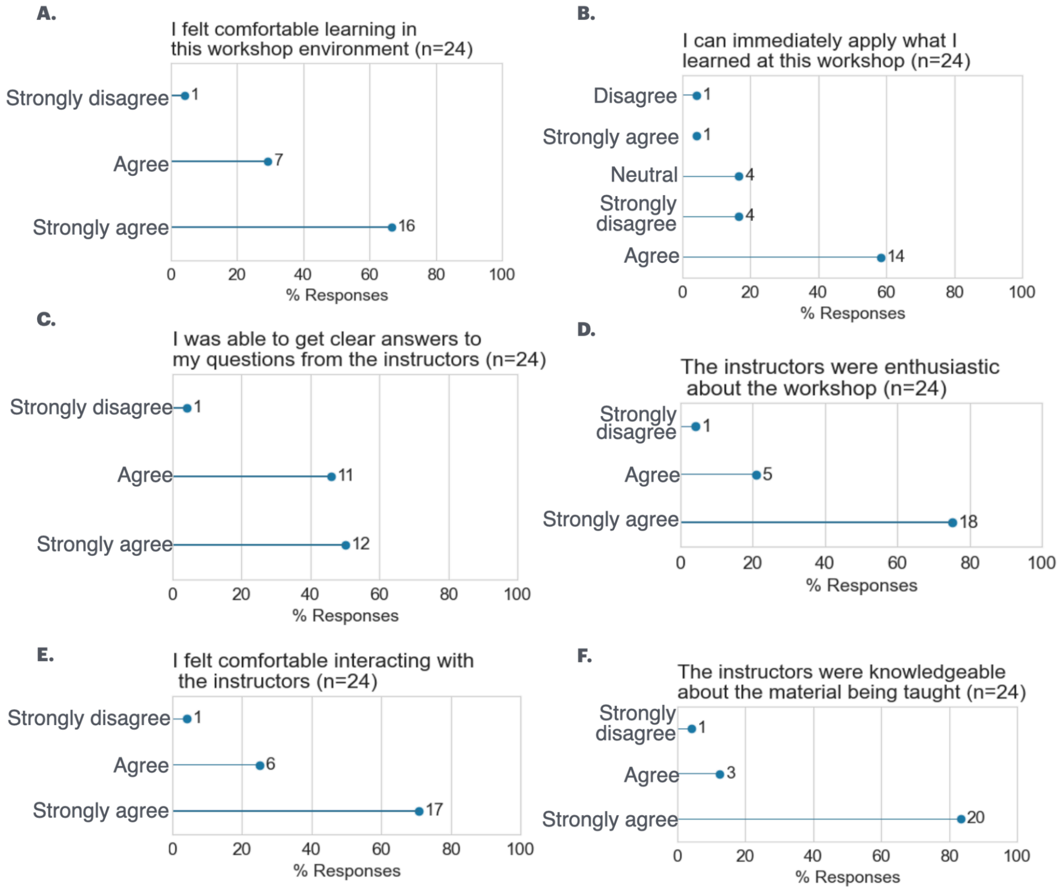


Figure 10 - Rating of participant's agreement with the following statements: **A** - I felt comfortable learning in this workshop environment; **B** - I felt comfortable interacting with the instructors (n = 22); **C** - The instructors were enthusiastic about the workshop (n = 22); **D** - The instructors were knowledgeable about the material being taught (n = 22); **E** - I was able to get clear answers to my questions from the instructors (n = 22); **F** - I can immediately apply what I learned at this workshop (n = 22).

This instructor's excellence, coupled with the optimum learning environment, culminated in high recommendation likelihood scores (Figure 11), with ~92% providing ratings of 8 or higher (respondents who chose "8", “9” and “10” were 4, 2 and 16 correspondingly). This suggests the workshop achieved its educational objectives for the vast majority of attendees, creating a replicable model for future ML education in biological sciences.

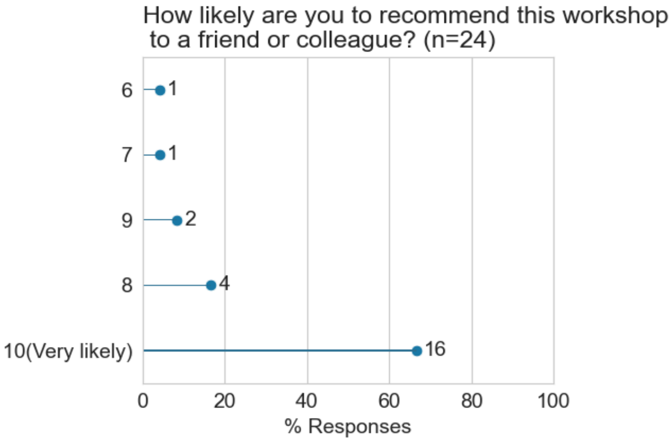


Figure 11 - Answers to the question: How likely would participants recommend this workshop to a friend or colleague? (n = 24); Rating interpretation - 1-3: No, 4-7: Maybe, 8-10: Yes; The answers from 0 to 5 are not shown in the plot as no responses were given in this range.

Overall feedback

The post-workshop survey collected participants' feedback on their learning experience using two questions: “Please list the major strengths of this workshop”, and “Please list the ways the workshop could be improved”. Recurring themes from the responses to these questions were identified, and the number of times each theme appeared was counted and recorded as “strength mentions” and “improvement mentions”. These counts provide a quantitative summary of participants' feedback regarding the workshops' strengths and areas for improvement. In total, 50 strength mentions were mentioned across all recurrent themes, compared to 18 improvement mentions (Figure 12), representing a ~2.7:1 ratio of strengths to areas for improvement. This indicates that participants found substantial value in the overall workshop experience.

Instructor quality and support, and Content Quality & Balance emerged as the program's greatest strengths, each receiving 13 mentions (Figure 12.A) highlighted the impact of knowledgeable, experienced instructors who can effectively bridge theoretical concepts with real-world applications. The strong performance in hands-on learning, receiving 9 mentions, indicated that the course successfully delivered on its core promise of practical, applicable ML education for biological data applications.

The improvement suggestions (i.e., ways the workshop could be improved), rather than indicating fundamental flaws, pointed toward opportunities to enhance an already solid educational foundation. The most frequently mentioned improvement area was hands-on Learning & Practice with 5 mentions (Figure 12.B). However, this category received 9 strength mentions (Figure 12.A), which reflects participants' desire for stronger or even more hands-on coding sessions rather than dissatisfaction with workshop offerings. The remaining improvement areas, each receiving only 1-3 mentions, focus primarily on workshop refinements and other operational aspects. This pattern suggests that participants are not seeking major structural changes but rather incremental improvements that would elevate their learning experience from good to exceptional, indicating a mature program ready for targeted enhancements rather than complete revision.

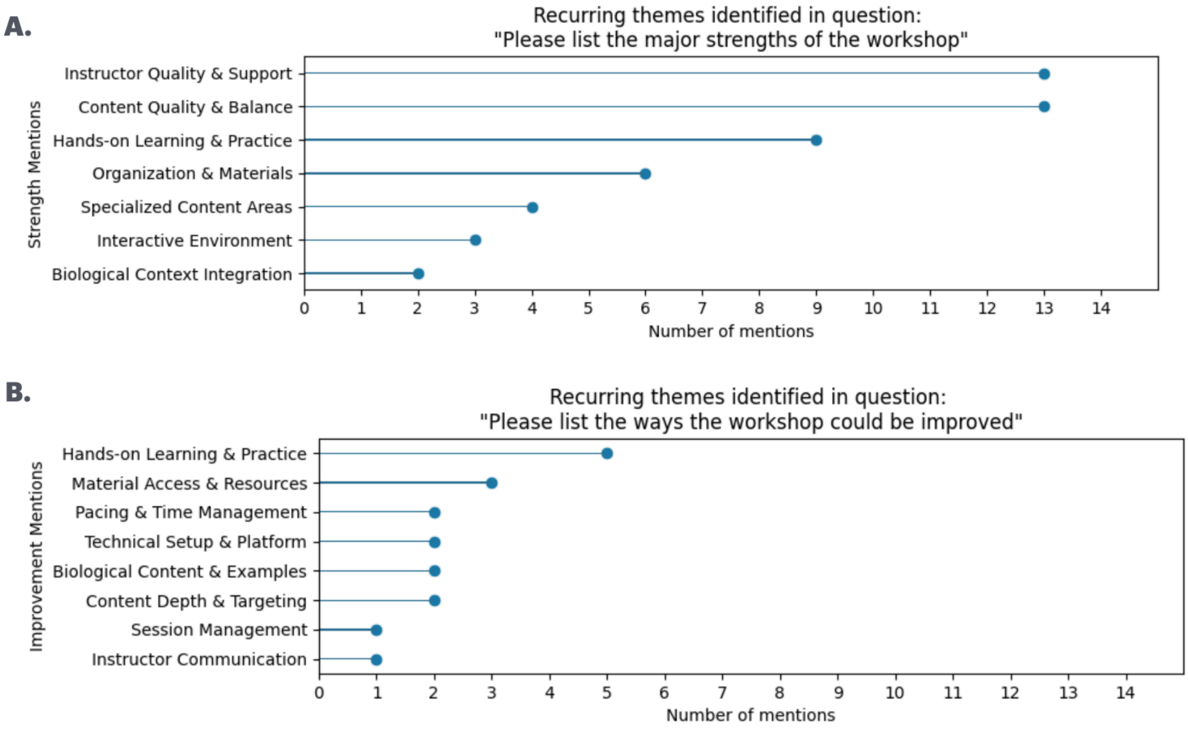


Figure 12 - Number of mentions providing a quantitative summary of strengths (A) and areas for improvement (B) of the workshop

Participants' self-reported improvement in skill and confidence

The pre- and post-workshop assessment revealed consistent and substantial improvements across all seven measured competency areas (as specified in the legend of Figure 13), demonstrating the workshop's effectiveness in building both technical skills and professional confidence among participants (Figure 13). Every assessed dimension showed positive gains, with participants progressing from moderate confidence levels (averaging ~0.3-1.0 on the pre-workshop scale) to strong confidence levels (averaging ~1.0-1.7 post-workshop). This uniform improvement pattern indicates that the workshop maintained high educational standards across different skill development areas, from foundational concepts like data reproducibility awareness to advanced applications such as custom programming capability.

The most significant improvements occurred in practical programming applications, particularly in participants' ability to write custom scripts for work-related problems and their overall confidence in using programming software for data analysis. These gains are especially meaningful given that participants already possessed considerable technical backgrounds, suggesting the workshop effectively advanced existing competencies rather than merely introducing basic concepts. These overall improvements in skill and confidence levels indicate that the workshop ultimately positioned participants for better integration of ML approaches into their biological research workflows.

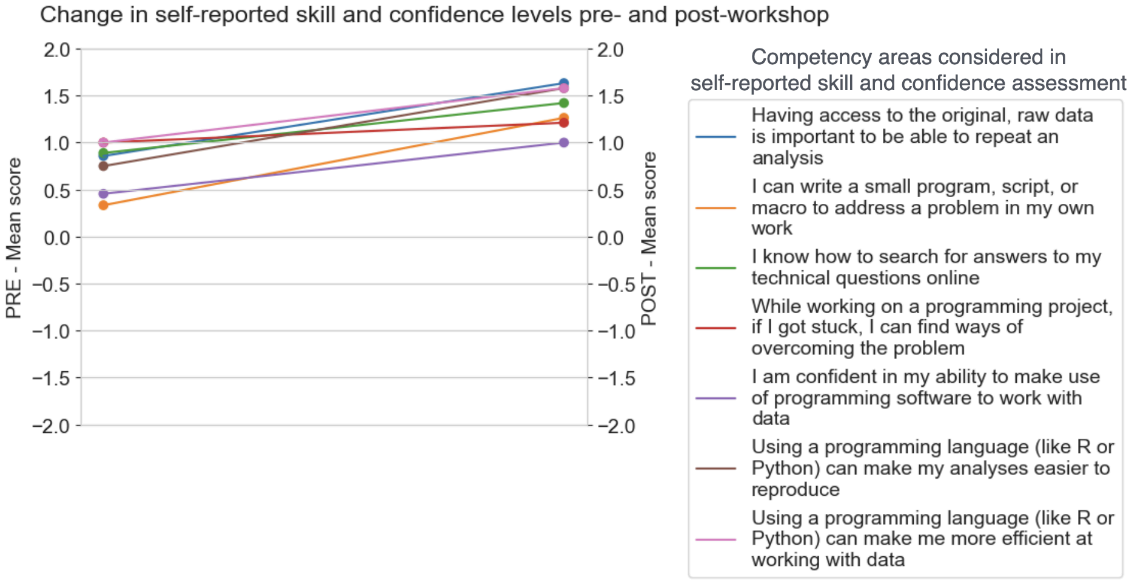


Figure 13 - Self-reported skill and confidence levels of participants based on pre- and post-workshop survey responses. Participants who submitted both pre- and post-workshop surveys (i.e., unique identifiers in both surveys; n = 18) were scored on a predefined scheme (-2: Strongly disagree, -1: Disagree, 0: Neutral, 1: Agree, 2: Strongly Agree) and the mean score was calculated for each category described in the legend

The tangible improvement in self-reported skills among participants was not only reflected in the survey responses but was also prominently displayed through their public professional profiles. Following the successful completion of the workshop (i.e., receiving the certificates), several participants chose to share their experiences on their LinkedIn feeds (Figure 14). These posts served as a powerful, organic testament to the value and impact of the workshop, demonstrating an authentic appreciation. This public endorsement highlighted the practical benefits gained, as individuals confidently showcased their newly acquired or enhanced abilities to their professional networks, further solidifying the workshop's efficacy.

LinkedIn post

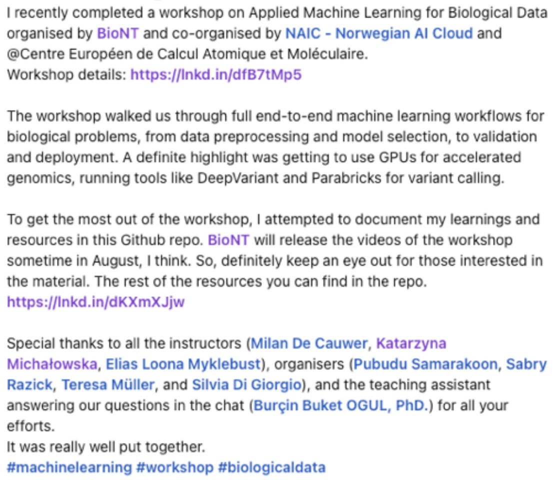

I'm happy to share that I've obtained a new certification: Applied Machine Learning for Biological Data from BioNT!

I found this course to be a great intro to machine learning for biological data, and will provide me with a solid foundation for further learning (thankfully all available afterward for review! Quite a bit to take in for a week).

What stands out the most to me is that such a high quality course was was free! I honestly cannot believe it. I'd like to pass on a special thanks to the organisers (Pubudu Samarakoon), speakers and organisations who fund BioNT. What a resource, I'll look out for more of their courses in future.

Celebrating a New Certification

LinkedIn post

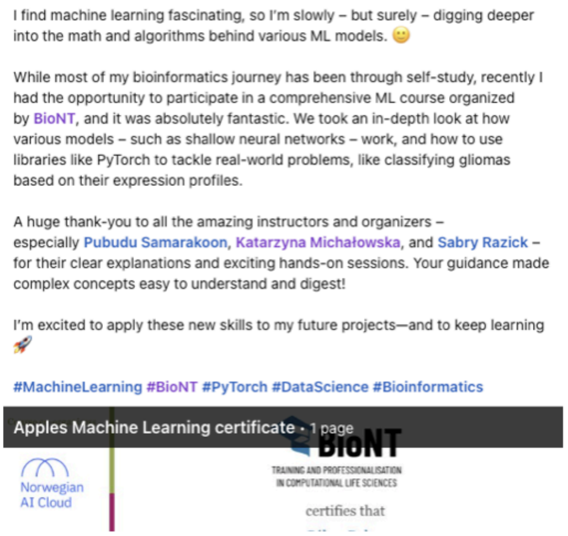

I recently completed a workshop on Applied Machine Learning for Biological Data organised by BioNT and co-organised by NAIC - Norwegian AI Cloud and @Centre Européen de Calcul Atomique et Moléculaire.
Workshop details: https://lnkd.in/dfB7tMp5

The workshop walked us through full end-to-end machine learning workflows for biological problems, from data preprocessing and model selection, to validation and deployment. A definite highlight was getting to use GPUs for accelerated genomics, running tools like DeepVariant and Parabricks for variant calling.

To get the most out of the workshop, I attempted to document my learnings and resources in this Github repo. BioNT will release the videos of the workshop sometime in August, I think. So, definitely keep an eye out for those interested in the material. The rest of the resources you can find in the repo.
https://lnkd.in/dKXmXJjw

Special thanks to all the instructors (Milan De Cauwer, Katarzyna Michałowska, Elias Loona Myklebust), organisers (Pubudu Samarakoon, Sabry Razick, Teresa Müller, and Silvia Di Giorgio), and the teaching assistant answering our questions in the chat (Burçin Buket OGUL, PhD.) for all your efforts.
It was really well put together.
#machinelearning #workshop #biologicaldata

LinkedIn post


I find machine learning fascinating, so I'm slowly – but surely – digging deeper into the math and algorithms behind various ML models. 🙂

While most of my bioinformatics journey has been through self-study, recently I had the opportunity to participate in a comprehensive ML course organized by BioNT, and it was absolutely fantastic. We took an in-depth look at how various models – such as shallow neural networks – work, and how to use libraries like PyTorch to tackle real-world problems, like classifying gliomas based on their expression profiles.

A huge thank-you to all the amazing instructors and organizers – especially Pubudu Samarakoon, Katarzyna Michałowska, and Sabry Razick – for their clear explanations and exciting hands-on sessions. Your guidance made complex concepts easy to understand and digest!

I'm excited to apply these new skills to my future projects—and to keep learning 🚀

#MachineLearning #BioNT #PyTorch #DataScience #Bioinformatics

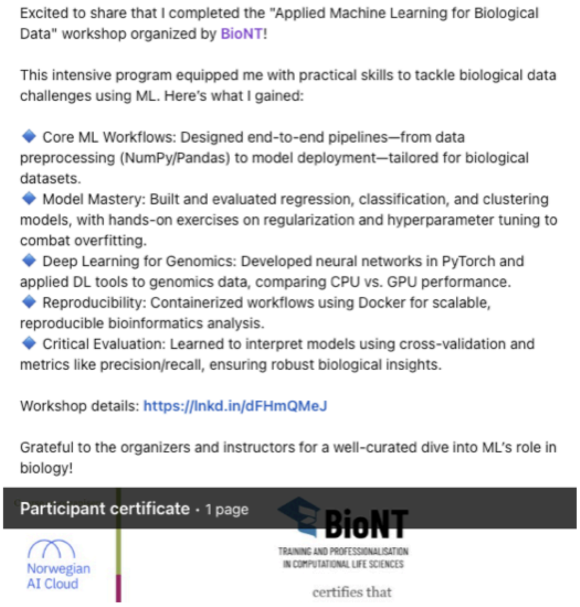
Apples Machine Learning certificate · 1 page

BioNT
TRAINING AND PROFESSIONALISATION IN COMPUTATIONAL LIFE SCIENCES

Norwegian AI Cloud

certifies that

LinkedIn post


Excited to share that I completed the "Applied Machine Learning for Biological Data" workshop organized by BioNT!

This intensive program equipped me with practical skills to tackle biological data challenges using ML. Here's what I gained:

🔷 Core ML Workflows: Designed end-to-end pipelines—from data preprocessing (NumPy/Pandas) to model deployment—tailored for biological datasets.
🔷 Model Mastery: Built and evaluated regression, classification, and clustering models, with hands-on exercises on regularization and hyperparameter tuning to combat overfitting.
🔷 Deep Learning for Genomics: Developed neural networks in PyTorch and applied DL tools to genomics data, comparing CPU vs. GPU performance.
🔷 Reproducibility: Containerized workflows using Docker for scalable, reproducible bioinformatics analysis.
🔷 Critical Evaluation: Learned to interpret models using cross-validation and metrics like precision/recall, ensuring robust biological insights.

Workshop details: https://lnkd.in/dFHmQMeJ

Grateful to the organizers and instructors for a well-curated dive into ML's role in biology!

Participant certificate · 1 page

BioNT
TRAINING AND PROFESSIONALISATION IN COMPUTATIONAL LIFE SCIENCES

Norwegian AI Cloud

certifies that

Figure 14 - Public posts of participants on their LinkedIn profiles endorsing the workshop (names of the participants were not shown to maintain anonymity)

Conclusion

The 8th BioNT workshop, "Applied Machine Learning for Biological Data," successfully concluded between May 27th and 28th and June 2nd-6th, 2025. To complement the main workshop, an industry meetup event titled "AI in Biomedical Data - Industry Seminar: Bridging Innovation & Opportunity" was organised on June 7th. The event was designed as a hybrid format with live streaming to attract a wider audience. This meetup event effectively promoted the workshop to industry employees and academics interested in transitioning to industry roles, particularly within SMEs. The success of this outreach strategy is evidenced by the survey responses discussed earlier in this report.

The setup for the 8th workshop benefited from the experience gathered during the first seven BioNT workshops and had additional attention to actively reach the audience, making sure the audience had a chance to fulfil the prerequisites.

This workshop saw a good engagement from pre-workshop events and through registration. There was an oversubscription and 40 participants were selected using a consistent criterion to project goals, where participants in Europe and seeking a job or were involved in an SME received priority.

The heightened interest, as evidenced by good participation numbers, can be attributed to more extensive advertising efforts, highlighting the effectiveness of increased partner engagement in promotions. Additionally, the broad appeal of the workshop, catering to researchers, bioinformaticians, data scientists, and professionals across academia and industry, further contributed to its success.

One of the distinguishing factors for the main workshop was its invited instructors, who regularly worked in the industry. Another important feature of the workshop was that it included industry-relevant, extensive hands-on sessions that followed the introduction of ML concepts and short ML exercises.

There were some significant challenges during this workshop. One of the anticipated risks in the project materialised for this event, similarly to previous BioNT workshops, which was the lack of trainers within the consortium to deliver some of the lectures. During the proposal, UO had planned to host this course together with SIMULA. Due to their departure from the project and the replacement partner not being able to deliver ML training, UO, as the main organiser, had to take mitigatory measures. One way was to recruit experienced lecturers from the field, as these lecturers from institutes not part of the BioNT consortium had to be paid by UO. This was possible thanks to a sponsorship from the Norwegian AI Cloud (NAIC). The other challenge was related to the planning of the specialised infrastructure needed for the course, where organisers faced delays in planning the VM equipment needed for the course and its respective funding. Therefore, to deliver a high-quality and timely training experience, UO proactively and successfully mitigated this risk by partnering with the NAIC, obtaining their support through a partial sponsorship.

Despite the mentioned risks and challenges, BioNT partners contributed heavily to this workshop to make it a success. As discussed in various sections of the report, participants consistently reported a comfortable and supportive learning environment, high satisfaction with the knowledgeable and enthusiastic instructors, and the immediate applicability of the course material. The workshop boosted participants' self-reported skill and confidence in several areas, including their ability to write custom scripts for work-related problems and their overall confidence in using programming software for data analysis. The high positive recommendation rate, coupled with feedback received during and after the event and positive LinkedIn posts, further confirms the workshop's success in meeting participant needs and expectations.

.