

Application: Predictive models evaluation & inspection in scikit-learn

Guillaume Lemaitre - g.lemaitre58@gmail.com
EOSS6: Essential Open Source Software for Science (Cycle 6)

Summary

ID: EOSS6-0000000551
Last submitted: Dec 4 2023 05:31 PM (CET)

1. Applicant Details

Completed - Oct 16 2023

1. Applicant Details

Complete the following information for the Applicant (required)

The information entered should be for the individual submitting the application who will act as the main person responsible for the application and as its point of contact. **To edit your name or email**, navigate to Account Information by clicking your name in the upper right corner. Please note that this person must remain the same between the LOI and full application (if selected).

Name: Guillaume Lemaitre

Email: g.lemaitre58@gmail.com

Add your home institution, company, or organization. This does not need to be the organization to which a grant would ultimately be awarded, if selected for funding.

Institution/Affiliation	scikit-learn consortium @ Inria Foundation
-------------------------	--

Add the institution, company, or organization that will be receiving the award. This may be the same as listed above, or different.

Award Organization	NumFOCUS
--------------------	----------

ORCID iD

Enter in format XXXX-XXXX-XXXX-XXXX. ORCID iDs are unique, digital identifiers that distinguish individual scientists and unambiguously connect their contributions to science over time and across changes of name, location, and institutional affiliation. ORCID iDs will be used to streamline reporting in our applications and grant reports to reduce the burden on grantees. For more information, please visit <https://orcid.org/register>. (Please contact us at sciencegrants@chanzuckerberg.com if you wish to opt out.)

0000-0002-0897-6791

2. Proposal Details

Completed - Oct 16 2023

2. Proposal Details

a. Proposal Title: Predictive models evaluation & inspection in scikit-learn

*Auto-filled; Maximum of 60 characters, including spaces. If you need to edit your proposal title, navigate to your application summary page; click on the three dots to the right of the application title; and select Rename from the dropdown menu. **Please note that you will not be able to make changes to the title of your application between the LOI and full proposal period.***

b. Amount Requested

Total budget amount requested in USD, including indirect costs; this number should be between \$100,000 USD and \$400,000 USD total costs over a two-year period. Enter whole numbers only (no dollar signs, commas, or cents).

400000

c. Proposal Summary/Scope of Work

Provide a short summary of the work being proposed (maximum of 500 words)

Scikit-learn is one of the fundamental open-source libraries for developing machine learning pipelines in both academic and industrial research. Its strength lies in providing a range of building blocks, allowing non-experts in applied mathematics to build machine learning pipelines within their domains of expertise. In addition to the consistent programming interface and a large range of statistical algorithms, the library comes with outstanding documentation to ease and democratize the use of machine learning in those domain-specific applications. This proposal aims to develop, enhance, and maintain the scikit-learn tools dedicated to evaluating, inspecting, and interpreting machine learning models.

When building a machine learning pipeline for a specific research problem, two key aspects are closely connected: (i) design the pipeline and (ii) assess, analyze, and inspect it. Researchers strive to identify the optimal pipeline, maximizing specific evaluation metrics, while also seeking at explaining the validity and rationale behind the pipeline's predictions. This is the corner stone to properly answering research questions. With this proposal we aim to improve and extend the available scikit-learn tools.

Scikit-learn provides building blocks for model evaluation and statistical analysis of results. Originally, this information was presented in a raw format and required expertise from scientists to create intuitive supports for outreach to peers and outsiders. Recently, the scikit-learn community developed displays to easily generate visual figures for communicating such results. However, these displays are still in their early development stages and do not leverage all available statistical analysis tools (i.e., cross-validation) from scikit-learn. Thus, we aim to expand these displays to use the right statistical tools and thus promote the adoption of best practices when reporting results. Additionally, we also intend to create new displays that are considered as "standards" during such analyses.

In the domain of model inspection, we aim to address several areas: (i) model inspection during training, (ii) enhancing user experience through interactive inspection, and (iii) model explainability. First, during the training of a pipeline, researchers are interested in monitoring the internal characteristics of the model and it is a not yet addressed long-standing issue in scikit-learn. We want to build upon some initial work by implementing a "callback" framework that allows users to track these internal parameters. Next, researchers commonly use interactive tools such as Jupyter Notebook to develop pipelines. Scikit-learn started some efforts to visually and interactively display pipelines in these environments. However, there is room for improvement in terms of user interaction and accessibility. Finally, as scikit-learn is widely used as a reference package, it is crucial to improve the section of the library dedicated to model explainability. We aim to improve the documentation and user experience with the

existing explainability tools, making sure that they use the appropriate tool for their use cases.

On top of all these items, we intend to continue working on the general maintenance of the project, addressing bug reports and performance regressions. As a community-driven project, we want to dedicate time reviewing external contributions.

d. Value to Biomedical Users

Described the expected value of the proposed work to the biomedical research community (maximum of 250 words)

Scikit-learn is a fundamental library for researchers and practitioners in the biomedical domain who aspire to develop machine learning pipelines. It provides building blocks for various stages, ranging from data preprocessing to machine learning algorithms (e.g., classification, regression). Once the machine learning pipeline is designed, scikit-learn also provides tools to optimize, assess, and inspect it. A search in Nature (<https://www.nature.com/search?q=scikit-learn>) returns around 3,000 scientific papers, with a majority for biomedical research.

The various lines of work outlined in the earlier proposal will particularly impact this model evaluation and inspection stages. Specifically, our objectives include enhancing the robustness and intuitiveness of these steps for our users, creating dissemination materials to share with their community, and ensuring the adoption of methodological best practices out-of-the-box.

As a specific example, we aim to simplify the process of plotting a Receiver Operating Characteristic (ROC) curve resulting from a cross-validation analysis. Our goal is to condense the code into just a few lines of Python, eliminating unnecessary boilerplate while ensuring the correctness of the methodology.

e. Open Source Software Projects

Number of software projects are involved in your proposal (maximum of five):

1

Complete the table with the following information for each software project. If there is no homepage URL, re-enter the main code repository URL.

	Software project name	Main code repository URL	Homepage URL
1	scikit-learn	https://github.com/scikit-learn/scikit-learn	https://scikit-learn.org/

f. Landscape Analysis

Briefly describe the other software tools (either proprietary or open source) that the audience for this proposal primarily uses. How do the software project(s) in this proposal compare to these other tools in terms of user base size, usage, and maturity? How do existing tools and the project(s) in this proposal interact? (maximum of 250 words)

Scikit-learn is one of the oldest and widely used libraries for machine learning in Python. It accounts for around 2,700 contributors and is used in approximately 600,000 GitHub repositories, with around 13,500 software packages depending on it.

Among the CZI-funded projects, several are directly dependent on scikit-learn, including CellProfiler, DeepLabCut, FastSurfer, ilastik, MDAnalysis, MNE-Python, NiLearn, NiPreps, Orange Data Mining, QIIME, and ScanPy.

Scikit-learn relies directly on fundamental Python libraries, including NumPy, SciPy, pandas, and Matplotlib, which are also supported by CZI funding. NumPy and SciPy are used to build machine learning algorithms, while Matplotlib is employed to create visual artifacts for model evaluation and inspection, notably. Ongoing work, detailed at <https://labs.quansight.org/blog/array-api-support-scikit-learn>, aims to enhance interoperability with other array libraries like CuPy or PyTorch to leverage Graphical Processing Units (GPUs) for accelerated computing.

Scientists also use deep learning frameworks such as TensorFlow or PyTorch, as well as gradient boosting libraries like XGBoost, LightGBM, or CatBoost. While deep learning is beyond the scope of scikit-learn, gradient-boosting-based libraries still rely on building blocks provided by scikit-learn for intermediate stages of the machine learning pipeline.

For more examples of libraries that utilize scikit-learn, please refer to https://scikit-learn.org/stable/related_projects.html.

g. Category

Choose the two categories that best describe the software project(s) audience.

	Category
Category 1	Machine learning and data analysis
Category 2	Visualization

h. Previous Funding

Have you ever received grant funding from CZI, the Wellcome Trust, or the Kavli Foundation? Select Yes or No.

Yes

Please check the box(es) of the organization(s) from which you received funding.

Responses Selected:

Chan Zuckerberg Initiative

Did you previously apply for funding under the CZI EOSS program? Select Yes or No.

Yes

Have you previously received funding under the CZI EOSS program? If yes, please list your application ID in the format EOSS1-0000000001.

Responses Selected:

Yes, application ID:: EOSS4-0000000223

3. Terms and Conditions

Completed - Oct 16 2023

3. Terms and Conditions

Terms and Conditions

Please carefully read the below terms and conditions regarding grant policies and personal data.

Grant Policies

Funded applications will be subject to various grant conditions and policies. **Submission to this program, as well as checking the box below, will imply that your organization agrees to and will be able to comply with these conditions.** Funder specific policies are linked below:

- [CZI Grant Policies](#)
- [Wellcome Grant conditions & grant funding policies](#)
- While the Kavli Foundation does not have a specific grant policies document, if you have questions related to Kavli grant conditions, please contact science@kavlifoundation.org.

Responses Selected:

I understand and acknowledge the grant policies and conditions

Application and Personal Data

By submitting your application, you agree to share all submitted application data (i.e. name(s), contact details, role, professional details, organization, details of your proposal, ORCID iD) and sharing these personal data with the Wellcome Trust and Kavli Foundation (in addition to CZI) for the purpose of administering, managing and evaluating your application, as well as for assessing the effectiveness of our grants program. In addition, if you choose to, you can voluntarily provide demographic data in the following section of the application. If you choose to provide the data, check the box in section b.2. below to consent to CZI's data privacy and sharing policy. The demographic data / diversity data will be aggregated and anonymized and this anonymized data will be shared with the Wellcome Trust and The Kavli Foundation for diversity monitoring purposes. Applications and reviews will be subject to and processed in accordance with the privacy policies for all three organizations:

- [Wellcome Grants Privacy and Confidentiality](#)
- [Kavli Foundation Privacy Policy](#)
- [Chan Zuckerberg Initiative Privacy Policy](#)

Responses Selected:

Check the box to acknowledge that you have read and understand the data privacy and sharing policy and consent to CZI sharing your LOI application data and subsequent full application data (if applicable) with the funders affiliated with this grant program (the Kavli Foundation and the Wellcome Trust).

Responses Selected:

Check the box to consent that you have read and understand the data privacy and sharing policy and consent to CZI collecting your optionally provided demographic / diversity data (as set out above), which will be aggregated and anonymized before being shared with the Wellcome Trust and The Kavli Foundation. Please note that providing any data is optional and all sections in the Equal Opportunity and Diversity section may be left blank. To withdraw your consent at any time please contact sciencegrants@chanzuckerberg.com.

Future Sharing

For unfunded proposals, we may share your proposal and reviews with other interested funders who may wish to pursue funding outside of the formal EOSS program. If you would like CZI to share your LOI proposal and subsequent full application data (if applicable) with other interested funders for potential funding, please check the “yes” box. We will notify the applicant and get consent before sharing. If you do not want your proposal to be shared, please select “no”.

Responses Selected:

yes

Continue onto the next section if you choose to provide optional demographic information. If you choose not to provide this data, you can submit your application.

4. Equal Opportunity & Diversity

Completed - Oct 16 2023

Equal Opportunity & Diversity

CZI Science supports the science and technology that will make it possible to cure, prevent, or manage all diseases by the end of this century. Different communities are affected by or experience disease in different ways. Moreover, due to systemic barriers, the scientific enterprise itself is not a place where all voices and talents thrive. We believe the strongest scientific teams — encompassing ourselves, our grantees, and our partners — incorporate a wide range of backgrounds, lived experiences, and perspectives that guide them to the most important unsolved problems. To enable our work, we incorporate diverse perspectives into our strategy and processes, and we also seek to empower community partners to engage in science.

We request demographic information associated with applications submitted to CZI in response to our open calls. This information helps us learn from the RFA process, as well as improve our strategies to help ensure members of underrepresented or marginalized groups in science are aware of and able to apply to CZI opportunities. **Please note that answering the questions below is voluntary, and receiving funding is not contingent on providing this information. Demographic information provided may be used in our grant-making process but will not be used as the sole or determinative factor in our grant funding decisions.** We may also publish aggregated data in various public forums, such as a website or blog. All responses will be shared only with limited personnel and service providers, who will use that information only for the purposes described in this paragraph.

If you have any additional questions about why we ask this, what we do with the data, or to share suggestions for improvement, please reach out to sciencegrants@chanzuckerberg.com.

eod_1

The information below may be entered for the Applicant, who will be submitting the application on behalf of the team. The categories listed below may not capture all possible identities; in the event that the categories do not accurately reflect your identities, please use the space provided to self-describe.

What is your race/ethnicity? (optional)

White (a person having origins in Europe, or otherwise identifies with this category)

What is the year of your last academic degree? (optional)

2016

What is your gender? (optional)

Man

Are you transgender? (optional)

No

Are you a member of the LGBTQIA+ community? (optional)

No

Do you have one or more disabilities and/or long term health conditions? (optional)

No

1. Applicant Details

Applicant Details

grantee_1

Please complete the following information for the Applicant (required):

The information entered should be for the individual submitting the application who will act as the main person responsible for the application and as its point of contact. Information about other Key Personnel on the proposal should be entered where requested in the Software Project Details part of the application. **Complete all fields in this task; all fields are required.** Some fields are auto-filled based on your answers in the LOI.

1. Name: Guillaume Lemaitre

2. Email: g.lemaitre58@gmail.com

3. Degrees (check all that apply)

Responses Selected:

PhD or equivalent

4. Institution/Affiliation	scikit-learn consortium @ Inria Foundation
5. Title/Position	Research Engineer
6. Country of Residence	France
7. Organization (Organization that would receive and distribute funding if a grant is successfully awarded. Note that this may be different from the Applicant's main affiliation/current employer.)	NumFOCUS

ORCID iD

Enter in format XXXX-XXXX-XXXX-XXXX. ORCID iDs are unique, digital identifiers that distinguish individual scientists and unambiguously connect their contributions to science over time and across changes of name, location, and institutional affiliation. ORCID iDs will be used to streamline reporting in our applications and grant reports to reduce the burden on grantees. For more information, please visit <https://orcid.org/register>. (Please contact us at sciencegrants@chanzuckerberg.com if you wish to opt out.)

0000-0002-0897-6791

2. Organization Details

Completed - Dec 1 2023

Organization Details

Please complete all requested information. The information entered should be for the organization that would be directly receiving and distributing funding if a grant is successfully awarded (e.g. academic institution, fiscal sponsor). Note that this may be different from the Applicant's main affiliation.

org_1

1. Type of Organization (required):

Fiscal sponsor

2a. Organization financial information (required):

Please provide the following information via a singular PDF upload. If the organization receiving the funds is a fiscal sponsor, please provide requested items 1-3 below for both the fiscally sponsored organization and the fiscal sponsor.

- Audited Financial Statements - please include the most recent fiscal year in addition to the previous fiscal year (2 years total); if audited financials are not available, please provide the unaudited version
- Organization Operating Budget - please include:
 1. Current Year Organization Budget; and
 2. A Forecasted Organization Budget which covers the duration of the potential grant [only if available]
- Organization Size - please provide the number of full-time employees employed by your organization, as of the current date
- Project list of funders: High-level only, total funding, confirmed vs. unconfirmed for the specific project for all years of the grant

[CZI Application Materials EOSS 6.pdf](#)

Filename: CZI Application Materials EOSS 6.pdf **Size:** 323.8 kB

3. Organization (required):

3a. Organization Name	NumFOCUS, Inc.
3b. Address	PO Box 90596
3c. City	78709 Austin
3d. State/Province	TX
3e. Country	United States
3f. Tax ID (9-digits; format: XX-XXXXXXX; enter 44-4444444 if not applicable)	45-4547709
3g. Organization website	https://www.numfocus.org

4. Organizational/Administrative Contact (required):

List the name and contact information for the administrative contact to discuss additional information needed, if selected for award.

4a. First Name	Miriam
4b. Last Name	Mueller
4c. Title/Position	Finance Director
4d. Email	finance@numfocus.org

5. Signing Official (required):

List the name and contact information for the person authorized to sign on behalf of your organization.

5a. First Name	Leah
5b. Last Name	Silen
5c. Title/Position	Executive Director
5d. Email	leah@numfocus.org

6. Press Contact / Public Relations Official (required):

List the name and contact information for the person to discuss press releases and media.

6a. First Name	Leah
6b. Last Name	Silen
6c. Title/Position	Executive Director
6d. Email	leah@numfocus.org

INSTITUTIONAL APPROVALS

Please note that there is no institutional approval form or agreement required from the Kavli Foundation.

7. CZI Institutional Approval Form (required):

Upload as a single PDF. This [form](#) should be reviewed and signed by a person authorized to sign on behalf of your organization agreeing to the stated institutional and investigator requirements and commitments on data, resource sharing, and publication policies, as well as endorsing/verifying your application materials and confirming their ability to receive funding for the proposal. In the event of an award, all funds will be awarded to the applicant institution as the prime institution, and the applicant institution will be responsible for ensuring compliance of all of the terms, including compliance of all partners/subcontract institutions. **These policies are non-negotiable so this form should only be signed if the organization is able to comply with the terms as stated.** While CZI does not require sign-off by all of your partner institutions, please refer to what your institution requires. **Note: digital signatures are permitted as long as the document is not encrypted or password-protected.**

[Cycle 6 Institutional Approval Form - signed.pdf](#)

Filename: Cycle 6 Institutional Approval Form - signed.pdf **Size:** 436.4 kB

8. Wellcome Grant Conditions (required):

By checking this box, I acknowledge that my organization has read and understands the [Wellcome Grant conditions](#).

Responses Selected:

My organization has read and understands these conditions.

3. Proposal Details

Completed - Dec 4 2023

Proposal Details

Please complete the following proposal information. All sections are required.

1. Proposal Title: Predictive models evaluation & inspection in scikit-learn

2. Previous Funding

Did you previously apply for funding for this or a related proposal under the CZI EOSS program?

Yes

Have you previously received funding for this proposal under the CZI EOSS program?

No

3. Proposal Purpose:

Describe the purpose of the proposal in one sentence (maximum of 200 characters including spaces). Example: To develop a comprehensive, validated atlas of the human kidney at single-cell resolution open to the entire scientific and clinical community.

o sustain the scikit-learn library by maintaining, improving, and extending it, notably in the domain of predictive model evaluation and inspection.

4. Amount Requested:

Enter the amount requested per year between \$50,000 USD and \$200,000 USD per year, including indirect costs), as well as the total budget requested for all years in USD (between \$100,000 USD and \$400,000 USD total, including indirect costs). These numbers should match those described in the Budget Description to follow. Enter whole numbers only (no dollar signs, commas, or cents)

Year One	191954
Year Two	197244
Total All Years	389198

5. Proposal Summary:

Provide a short summary of the application (maximum of 500 words) (auto-filled from LOI; update if needed)

Scikit-learn is one of the fundamental open-source libraries for developing machine learning pipelines in both academic and industrial research. Its strength lies in providing a range of building blocks, allowing non-experts in applied mathematics to build machine learning pipelines within their domains of expertise. In addition to the consistent programming interface and a large range of statistical algorithms, the library comes with outstanding documentation to ease and democratize the use of machine learning in those domain-specific applications. This proposal aims to develop, enhance, and maintain the scikit-learn tools dedicated to evaluating, inspecting, and interpreting machine learning models. When building a machine learning pipeline for a specific research problem, two key aspects are closely connected: (i) design the pipeline and (ii) assess, analyze, and inspect it. Researchers strive to identify the optimal pipeline, maximizing specific evaluation metrics, while also seeking at explaining the validity and rationale behind the pipeline's predictions. This is the corner stone to properly answering research questions. With this proposal we aim to improve and extend the available scikit-learn tools. Scikit-learn provides building blocks for model evaluation and statistical analysis of results. Originally, this information was presented in a raw format and required expertise from scientists to create intuitive supports for outreach to peers and outsiders. Recently, the scikit-learn community developed displays to easily generate visual figures for communicating such results. However, these displays are still in their early development stages and do not leverage all available statistical analysis tools (i.e., cross-validation) from scikit-learn. Thus, we aim to expand these displays to use the right statistical tools and thus promote the adoption of best practices when reporting results. Additionally, we also intend to create new displays that are considered as "standards" during such analyses. In the domain of model inspection, we aim to address several areas: (i) model inspection during training, (ii) enhancing user experience through interactive inspection, and (iii) model explainability. First, during the training of a pipeline, researchers are interested in monitoring the internal characteristics of the model and it is a not yet addressed long-standing issue in scikit-learn. We want to build upon some initial work by implementing a "callback" framework that allows users to track these internal parameters. Next, researchers commonly use interactive tools such as Jupyter Notebook to develop pipelines. Scikit-learn started some efforts to visually and interactively display pipelines in these environments. However, there is room for improvement in terms of user interaction and accessibility. Finally, as scikit-learn is widely used as a reference package, it is crucial to improve the section of the library dedicated to model explainability. We aim to improve the documentation and user experience with the existing explainability tools, making sure that they use the appropriate tool for their use cases. On top of all these items, we intend to continue working on the general maintenance of the project, addressing bug reports and performance regressions. As a community-driven project, we want to dedicate time reviewing external contributions.

6. Work Plan:

A description of the proposed work for which funding is being requested, including resources the applicants will provide that are not part of the requested funding. For software development-related work (e.g., engineering, product design, user research), specify how the work fits into the existing software project roadmap. For community outreach related activities (e.g., sprints, training), specify how these activities will be organized, the target audience, and expected outcomes (maximum of 750 words)

The scikit-learn community has a public roadmap [1] that is currently under revision for updating. While it is not yet publicly available, the areas around predictive model evaluation and inspection are reported as highly important. Thus, the items earlier presented in the “Proposal Summary” are part of this community roadmap. In the following, we present in more detail the concrete scikit-learn enhancements intended by this proposal.

We plan to extend the capabilities of the scikit-learn displays. These displays generate visual reports used to evaluate and inspect predictive models. Primary work has been done, but the current displays do not leverage cross-validation and thus lack reporting information related to uncertainty, which is important when evaluating and inspecting predictive models. To achieve this milestone, the evaluation functionalities of scikit-learn (e.g., per-fold indices or per-fold predictions) need to be extended and then leveraged by the displays. To further improve model evaluation, we plan to create additional displays that are currently missing but useful when it comes to model evaluation.

Then, we want to improve the user experience when it comes to model inspection.

A long-standing feature request is about getting information during the training phase of a predictive model. In this regard, we intend to develop a callback framework. We recently prototyped the foundations of this framework and we intend to (i) finalize this framework, (ii) further develop different types of callbacks, and (iii) extend all scikit-learn estimators so that they use these callbacks. Such a framework should, of course, be made available to third-party libraries compatible with scikit-learn.

Another type of model inspection is related to interactive model development: when developing predictive models in an environment such as Jupyter notebooks, users benefit from a visual representation of the model that they design. Scikit-learn provides a basic visual representation of the machine learning pipeline. However, those diagrams are quite limited and should be extended. Concretely, we want to provide representation to visualize models’ parameters, as well as information regarding the propagation of feature names. As for the callback framework, we want to make this feature accessible and extensible to third-party libraries so that they can be integrated within scikit-learn components.

Finally, we want to lay some groundwork regarding model explainability. Scikit-learn provides a couple of techniques in this category. However, we would like to come up with a unified API to ease the interaction between these tools and our users. We, therefore, intend to pursue the work launched in the following scikit-learn enhancement proposal [2]. In addition to a new API, we will also investigate the available techniques that should be considered state-of-the-art and could be included in scikit-learn.

While the previous items are intended to extend scikit-learn, we also want to dedicate extensive groundwork to the maintenance of scikit-learn. At the time of writing, the issue tracker has around 1,600 issues, and approximately 600 pull requests are made to the project. With the support of the CZI Cycle 6, the project succeeded in significantly reducing the number of pull requests. We would like to further continue this work by supporting contributors providing bug reports and bug fixes. Reviewing pull requests and quickly giving feedback to contributors helps ensure the project's sustainability and enriches the contributor experience. The project receives many pull requests and is constrained by maintainer time to review these indispensable contributions. Maintainers who do this work are a mixture of volunteers and some paid effort from Inria and Quansight.

To deliver this technical work while improving the repository's inclusiveness, we plan to scale our scikit-learn open-source mentoring program. The idea is to replicate at scale what we did with one NumFocus Grant and Stefanie Senger [3]. We want to offer 6-month grants for mentees keen on contributing to the package. Each mentee is paired with an experienced core contributor mentor, ensuring personalized guidance. Through structured onboarding, guided contributions aligned with the technical work packages described earlier, and skill development initiatives, mentees immerse themselves in the scikit-learn community, actively engaging in their contributor experience. Mentees showcase their achievements, transitioning into the contributor team. To complete this program, we plan to hold two 5-day sprints to boost the deliverables of the technical roadmap.

In terms of timing, 0.55 FTE of Lucy Liu will be dedicated to the general maintenance and development of displays and the unified API around model feature importance. The different mentees will dedicate 1 FTE to the different topics mentioned earlier.

[1] <https://scikit-learn.org/stable/roadmap.html>

[2] https://github.com/scikit-learn/enhancement_proposals/pull/86

[3] <https://blog.scikit-learn.org/diversity/mentoring/>

7. Milestones and Deliverables:

List expected milestones and deliverables, and their expected timeline. Be specific and include where possible any goals for metrics the software project(s) are expected to reach upon completion of the grant. Please use a third-person voice (maximum of 500 words).

The following milestones are expected for the items discussed in the “Work Plan”:

- Extend displays (3 release cycles)
 - 1 release cycle to extend the scikit-learn API
 - 1 release cycle to leverage the new API in the displays
 - 1 release cycle to develop a new type of displays
- Develop callbacks (3 release cycles)
 - 1 release cycle for the framework implementation
 - 1 release cycle to implement different types of callbacks
 - 1 release cycle to make the framework available to third-party libraries
- Extend interactive HTML diagram (3 release cycles)
 - 1 release cycle to improve the current estimator HTML diagram
 - 2 release cycles to implement new types of interactive diagrams for parameters and feature names
- Unified API for explainability (4 release cycles)
 - 1 release cycle to finalize the scikit-learn enhancement proposal
 - 2 release cycles to implement the new API
 - 1 release cycle to implement state-of-the-art techniques

The scikit-learn project releases with a cadence close to 6 months (corresponding to a release cycle). All the above deliverables include code implementation, tests, API documentation, user guide documentation, and examples. All durations are subject to change depending on maintainer availability to review them.

Providing precise milestones for the maintenance work is more complex due to the variance in time in issues and pull requests. Nevertheless, the objective remains focused on reducing the number of issues and pull requests currently present in the backlog by addressing and reviewing them faster than contributors create them.

For the mentorship program, the goal is to provide two 6-month grants every 6 months throughout the 2-year

duration of the CZI program.

Regarding the community goal, the objective is to grant Lucy Liu access to the core developer team. For the mentees in the mentorship program, the aim is to facilitate their access to the recurring contributor team of the project.

8. Existing Support:

List active and recently completed (previous two calendar years) financial or in-kind support for the software project(s), including duration, total costs in USD, and source of funding. Include any previous funding for these software projects received from CZI, Wellcome, and/or Kavli outside of the EOSS program (maximum of 250 words).

The following grants provided financial support for the scikit-learn project:

- Maintenance & Extension of Scikit-learn: Machine Learning in Python. CZI Cycle 6. 2021-2023. PI: Thomas J. Fan.
- NASA ROSES E7 grant.

Currently the following people are sponsored:

- scikit-learn consortium @ Inria Foundation: Arturo Amor, François Goupil, Guillaume Lemaître, Jérémie du Boisberranger, and Olivier Grisel. They are full-time funded to work on scikit-learn development and scikit-learn community management.
- Quansight Labs: 0.25 FTE for Lucy Liu.
- Nvidia: 0.5 FTE for Tim Head.

9. Landscape Analysis:

Describe the other software tools (either proprietary or open source) that the audience for this proposal primarily uses. How do the software project(s) in this proposal compare to these other tools in terms of user base size, usage, and maturity? How do existing tools and the project(s) in this proposal interact? (maximum of 250 words). (auto-filled from LOI; update if needed)

Scikit-learn is one of the oldest and widely used libraries for machine learning in Python. It accounts for around 2,700 contributors and is used in approximately 600,000 GitHub repositories, with around 13,500 software packages depending on it. Among the CZI-funded projects, several are directly dependent on scikit-learn, including CellProfiler, DeepLabCut, FastSurfer, ilastik, MDAnalysis, MNE-Python, NiLearn, NiPreps, Orange Data Mining, QIIME, and ScanPy. Scikit-learn relies directly on fundamental Python libraries, including NumPy, SciPy, pandas, and Matplotlib, which are also supported by CZI funding. NumPy and SciPy are used to build machine learning algorithms, while Matplotlib is employed to create visual artifacts for model evaluation and inspection, notably. Ongoing work, detailed at <https://labs.quansight.org/blog/array-api-support-scikit-learn>, aims to enhance interoperability with other array libraries like CuPy or PyTorch to leverage Graphical Processing Units (GPUs) for accelerated computing. Scientists also use deep learning frameworks such as TensorFlow or PyTorch, as well as gradient boosting libraries like XGBoost, LightGBM, or CatBoost. While deep learning is beyond the scope of scikit-learn, gradient-boosting-based libraries still rely on building blocks provided by scikit-learn for intermediate stages of the machine learning pipeline. For more examples of libraries that utilize scikit-learn, please refer to https://scikit-learn.org/stable/related_projects.html.

10. Value to Biomedical Users:

*Describe the expected value of the proposed work to the biomedical research community (maximum of 250 words).
(auto-filled from LOI; update if needed)*

Scikit-learn is a fundamental library for researchers and practitioners in the biomedical domain who aspire to develop machine learning pipelines. It provides building blocks for various stages, ranging from data preprocessing to machine learning algorithms (e.g., classification, regression). Once the machine learning pipeline is designed, scikit-learn also provides tools to optimize, assess, and inspect it. A search in Nature (<https://www.nature.com/search?q=scikit-learn>) returns around 3,000 scientific papers, with a majority for biomedical research. The various lines of work outlined in the earlier proposal will particularly impact this model evaluation and inspection stages. Specifically, our objectives include enhancing the robustness and intuitiveness of these steps for our users, creating dissemination materials to share with their community, and ensuring the adoption of methodological best practices out-of-the-box. As a specific example, we aim to simplify the process of plotting a Receiver Operating Characteristic (ROC) curve resulting from a cross-validation analysis. Our goal is to condense the code into just a few lines of Python, eliminating unnecessary boilerplate while ensuring the correctness of the methodology.

11. Category:

Choose the two categories that best describe the software project(s) audience

Category 1	Machine learning and data analysis
Category 2	Visualization

12. Diversity, Equity, and Inclusion Statement:

Advancing DEI is a core value for this program, and we are requesting information on your efforts in this area. Describe any efforts the software project(s) named in this proposal have undertaken to increase diversity, equity, and inclusion with respect to their contributors and audience. Please see [examples](#) from applications funded in previous cycles (maximum of 250 words)

The scikit-learn project and its community strive for increasing diversity, equity, and inclusion with respect to their contributors and audience. Below, we outline specific efforts pursued over several years.

The oldest initiative involves coding sprints to assist first-time contributors in engaging with open-source development. These events are co-organized with communities that reach out to underrepresented groups, thereby expanding our impact. Collaborations include partnerships with Data Umbrella, Women in Machine Learning and Data Science (WiMLDS), and PyLadies. These efforts have a considerable impact on the involved contributors and the scikit-learn community [1, 2].

These efforts notably lead to changes in the scikit-learn governance [3], aiming to better acknowledge all types of contributions, not solely focusing on code. Although ongoing work is necessary in this area, the emerging changes signal a positive direction for the community.

Building on the success of coding sprints, a subsequent effort is the mentorship program, as discussed in the “Work Plan.” The program involves providing a 6-month grant to onboard and mentor contributors to contribute to scikit-learn. To promote diversity, equity, and inclusion, we collaborate with PyLadies to reach underrepresented groups. This effort has been successful, as exemplified by Stefanie Senger’s contributions. This proposal seeks to scale up this initiative to replicate such success [4].

[1] <https://blog.dataumbrella.org/impact>

[2] https://www.youtube.com/watch?v=-ioFzLNo_E4&ab

[3] <https://scikit-learn.org/dev/governance.html>

[4] https://www.linkedin.com/posts/stefaniesenger_scikitlearn-internship-opensource-activity-7073765657196011520-yvBt

4. Optional Attachments

Incomplete

Attachments should be uploaded in a combined single PDF. This section can include figures, charts and tables, references for the proposal, or any additional material in support of the proposal (maximum of three pages). Uploading any additional information is optional.

5. CV of Applicant

Completed - Dec 1 2023

Upload in PDF format; include current and recent employment, education history, and references to any major publications, software contributions, or other relevant outputs (maximum of two pages)

[CV-CZI](#)

Filename: CV-CZI.pdf Size: 120.6 kB

6. Budget Description

Completed - Dec 1 2023

Upload in PDF format; budgets can be uploaded in a combined single PDF or one PDF for each software project; font must be 11 point or larger and margins must be at least one-half inch (top, bottom, left and right) for all pages (letter size required, one page per software project maximum).

- Description of the costs to be funded by this grant at a high level and in narrative or tabular form, outlining costs for personnel (including names, if known), supplies, equipment, travel, meetings/hackathons/sprints, subcontracts, other costs, and up to 15% indirect costs (excluding equipment and subcontracts).
- Indirect costs are limited to up to 15% of direct costs and are included within the annual budget total. Indirect costs may not be assessed on capital equipment or subcontracts, but subcontractors may include up to 15% indirect costs of their direct costs. **Non-charitable entities must include a clear allocation and explanation for any indirect costs included in a proposed budget.**
- Budget should be requested in US dollars.
- International grantees must use all grant funds exclusively for activities conducted outside the United States of America. Travel expenses to the United States (including round-trip tickets) should not be covered from the requested grant funds.
- Application budgets must reflect the actual needs of the proposal. The Funders will work closely with successful applicants to arrive at a mutually acceptable budget after review.

[CZI EOSS6 budget](#)

Filename: CZI_EOSS6_budget.pdf Size: 30.2 kB

7. Open Source Software Project(s) Details

Completed - Dec 1 2023

Open Source Software Project Details

Provide details and metrics for each open source software project that will be supported by the grant to help us assess its impact and quality.

How many software projects involved in your proposal (up to five)? If multiple software projects are involved, details must be entered for all of them where indicated below.

1

SOFTWARE PROJECT #1

SOFTWARE PROJECT #1: Details:

Complete the following table for Open Source Software Project #1 of your proposal. All URLs should be in the format <https://example.com> and only one primary link should be provided.

Software Project name (required)	scikit-learn
Main code repository (e.g. GitHub URL) (required)	https://github.com/scikit-learn/scikit-learn
Homepage URL (required)	https://scikit-learn.org/
Social media handles (if applicable)	https://twitter.com/scikit_learn
Do you or software project key personnel have commit rights to the code repositories for this software project? (required)	Yes
Short description of software project (200 words maximum) (required)	<p>The scikit-learn project is an open-source project written for Python, implementing state-of-the-art machine learning algorithms and utilities for real-world data analysis and prediction problems. It has become the cornerstone of applied machine learning and data science in academic and industrial research. One of the project's primary goals is to make machine learning techniques more accessible and easy to use. In addition, Scikit-learn aims to empower software engineers and researchers across disciplines without machine learning expertise to build models and gain insights into their data.</p>

SOFTWARE PROJECT #1: List of Key Personnel:

Key personnel are people involved in the software project who will be supported by the grant if the application is successful.

Complete the following for the key personnel on the open source software project #1 (up to 5) (required); **enter n/a if any field is not applicable**. Personnel to be hired that have not been identified at this time can be listed in the budget section. You may need to use the scroll bar at the bottom of the table to scroll right to view and to complete all fields. Alternatively, you can tab to move through and complete the fields. **To add another person/row (up to five), click the box at the end of the row.**

	First name	Last name	Email address	Current employer/ Affiliation	Role in project	Developer username if applicable (e.g., GitHub handle)	Country of Residence	Add another person/row
1	Liu	Lucy	lliu@quansight.com	Quansight , LLC	Software Engineer	lucyleeow	Australia	✕

SOFTWARE PROJECT #1: Metrics- Quality (required):

Complete for the open source software project #1.

1. SOFTWARE PROJECT #1 : What is the software project license?

Permissive license (e.g. BSD 3-Clause, MIT, Apache 2.0)

2. SOFTWARE PROJECT #1 : What is the main programming language?

Python

3. SOFTWARE PROJECT #1: Does the software project have a code of conduct?

https://github.com/scikit-learn/scikit-learn/blob/main/CODE_OF_CONDUCT.md

4. SOFTWARE PROJECT #1: Does the software project have end-user documentation?

https://scikit-learn.org/stable/user_guide.html

5. SOFTWARE PROJECT #1: Does the software project have an issue tracker?

<https://github.com/scikit-learn/scikit-learn/issues>

6. SOFTWARE PROJECT #1: Does the software project have a community engagement / Q&A forum (self-hosted, on Stack Exchange etc.)?

<https://github.com/scikit-learn/scikit-learn/discussions>

7. SOFTWARE PROJECT #1: Does the software project have contribution / coding guidelines?

<https://scikit-learn.org/stable/developers/contributing.html>

8. SOFTWARE PROJECT #1: Is there a corresponding package available in a package manager (PyPi, CRAN, etc.)?

<https://pypi.org/project/scikit-learn/>

9. SOFTWARE PROJECT #1: Does the software project support continuous integration for testing?

We use a combination of GitHub Action, Circle CI, Azure Pipeline, and Cirrus

SOFTWARE PROJECT #1: Metrics- Impact (optional):

Complete the following for the open source software project #1. **Providing metrics is optional and metrics can be approximate.** For each metric, please provide a source, clarify how the metric was computed, and/or provide any other comments. For monthly metrics, please provide data from the most recent month for which the corresponding metric is available.

1. SOFTWARE PROJECT #1: Complete the following table. List the number and explanation for each, if needed:

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project	83240	Based on Google Scholar for the paper: Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.
Monthly users, if applicable (based on one or more of the following: monthly downloads from websites, monthly downloads from package managers, monthly unique requests for updates, etc.)	43703789	Based on the statistic on PyPI: https://pypistats.org/packages/scikit-learn
Software projects that depend on the project	14227	https://github.com/scikit-learn/scikit-learn/network/dependents?dependent_type=PACKAGE
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow)	906000	Based on google analytics to the scikit-learn.org website

2. SOFTWARE PROJECT #1: List of software projects to which key personnel on this proposal are contributing.

Lucy Liu is contributing to the following open-source project: scikit-learn (<https://github.com/scikit-learn/scikit-learn>), sphinx-gallery (<https://github.com/sphinx-gallery/sphinx-gallery>), napari (<https://github.com/napari/napari>)

8. Terms and Conditions

Completed - Dec 1 2023

Terms and Conditions

Terms and Conditions

Please carefully read the below terms and conditions regarding grant policies and personal data.

Grant Policies

Funded applications will be subject to various grant conditions and policies. **Submission to this program, as well as checking the box below, will imply that your organization agrees to and will be able to comply with these conditions.** Funder specific policies are linked below:

- [CZI Grant Policies](#)
- [Wellcome Grant conditions & grant funding policies](#)
- While the Kavli Foundation does not have a specific grant policies document, if you have questions related to Kavli grant conditions, please contact science@kavlifoundation.org.

Responses Selected:

I understand and acknowledge the grant policies and conditions

Application and Personal Data

By submitting your application, you agree to share all submitted application data (i.e. name(s), contact details, role, professional details, organization, details of your proposal, ORCID iD) and sharing these personal data with the Wellcome Trust and Kavli Foundation (in addition to CZI) for the purpose of administering, managing and evaluating your application, as well as for assessing the effectiveness of our grants program. In addition, if you choose to, you can voluntarily provide demographic data in the following section of the application. If you choose to provide the data, check the box in section b.2. below to consent to CZI's data privacy and sharing policy. The demographic data / diversity data will be aggregated and anonymized and this anonymized data will be shared with the Wellcome Trust and The Kavli Foundation for diversity monitoring purposes. Applications and reviews will be subject to and processed in accordance with the privacy policies for all three organizations:

- [Wellcome Grants Privacy and Confidentiality](#)
- [Kavli Foundation Privacy Policy](#)
- [Chan Zuckerberg Initiative Privacy Policy](#)

Responses Selected:

Check the box to acknowledge that you have read and understand the data privacy and sharing policy and consent to CZI sharing your full application data with the funders affiliated with this grant program (the Kavli Foundation and the Wellcome Trust).

Responses Selected:

Check the box to consent that you have read and understand the data privacy and sharing policy and consent to CZI collecting your optionally provided demographic / diversity data (as set out above), which will be aggregated and anonymized before being shared with the Wellcome Trust and The Kavli Foundation. Please note that providing any data is optional and all sections in the Equal Opportunity and Diversity section may be left blank. To withdraw your consent at any time please contact sciencegrants@chanzuckerberg.com.

Future Sharing

For unfunded proposals, we may share your proposal and reviews with other interested funders who may wish to pursue funding outside of the formal EOSS program. If you would like CZI to share your full application data with other interested funders for potential funding, please check the “yes” box. We will notify the applicant and get consent before sharing. If you do not want your proposal to be shared, please select “no”.

Responses Selected:

yes

Continue onto the next section if you choose to provide optional demographic information. If you choose not to provide this data, you can submit your application.

9. Equal Opportunity & Diversity

Completed - Dec 1 2023

Equal Opportunity & Diversity

CZ Science supports the science and technology that will make it possible to cure, prevent, or manage all diseases by the end of this century. Different communities are affected by or experience disease in different ways. Moreover, due to systemic barriers, the scientific enterprise itself is not a place where all voices and talents thrive. We believe the strongest scientific teams — encompassing ourselves, our grantees, and our partners — incorporate a wide range of backgrounds, lived experiences, and perspectives that guide them to the most important unsolved problems. To enable our work, we incorporate diverse perspectives into our strategy and processes, and we also seek to empower community partners to engage in science.

We request demographic information associated with applications submitted to CZI in response to our open calls. This information helps us learn from the RFA process, as well as improve our strategies to help ensure members of underrepresented or marginalized groups in science are aware of and able to apply to CZI opportunities.

Please note that answering the questions below is voluntary, and receiving funding is not contingent on providing this information. Demographic

information provided may be used in our grant-making process but will not be used as the sole or determinative factor in our grant funding decisions. We may also publish aggregated data in various public forums, such as a website or blog. All responses will be shared only with limited personnel and service providers, who will use that information only for the purposes described in this paragraph.

If you have any additional questions about why we ask this, what we do with the data, or to share suggestions for improvement, please reach out to sciencegrants@chanzuckerberg.com.

The information below may be entered for the Applicant, who will be submitting the application on behalf of the team. The categories listed below may not capture all possible identities; in the event that the categories do not accurately reflect your identities, please use the space provided to self-describe.

eod_1

This section has been auto-filled based on data entered in the LOI for the Applicant. Please edit as needed. Please note that completing the below is voluntary, and demographic information will not be used to make final grant funding decisions.

	LOI response	Optional description
What is your race/ethnicity? (optional)	White (a person having origins in Europe, or otherwise identifies with this category)	

	LOI response	Optional description
What is the year of your last academic degree? (optional)	2016	

	LOI response	Optional description
What is your gender? (optional)	Man	

	LOI response	Optional description
Are you transgender? (optional)	No	

	LOI response	Optional description
Are you a member of the LGBTQ community?	No	

	LOI response	Optional description
Do you have one or more disabilities and/or long term health conditions? (optional)	No	

The Applicant may also provide aggregated information on the demographics of the Key Personnel listed in the Open Source Software Project Details section of the application if those personnel agree to share that information with CZI. The categories listed below may not capture all possible identities; in the event that the categories do not accurately reflect their identities, please use the space provided to describe.

How many of the Key Personnel self-identify as each of the following genders? (optional)

Woman	1
Man	(No response)
Non-binary/Third gender	(No response)
Prefer not to state	(No response)
Prefer to describe	(No response)

How many of the listed Key Personnel self-identify as one of the below race/ethnicities? (optional)

Please do not include requested information on a per person basis; we are looking for aggregated information.

Two or More Races (a person who identifies with more than one of the categories below)	(No response)
Black and/or African American (Afro-Caribbean, African American, Black and from an African nation, or a person who otherwise identifies with this category)	(No response)
Asian (a person having origins in the Far East, Southeast Asian, the Indian Subcontinent; this does not include Native Hawaiian or Other Pacific Islander)	(No response)
White (a person having origins in Europe, or otherwise identifies with this category)	(No response)
Hispanic or Latinx (a person having origins in Mexico, Central and/or South America, Puerto Rico, and/or Cuba)	(No response)
Middle Eastern or North African (a person having origins in or who maintain an affiliation to the Middle East or North African regions)	(No response)
Native Hawaiian or Other Pacific Islander (a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands)	(No response)
American Indian or Alaska Native (a person having origins in North and South America, including Central America, and who maintain tribal affiliation or community attachment)	(No response)
Prefer not to say	(No response)
Prefer to describe	(No response)

Guillaume Lemaître

Research Engineer

35 years old

Scikit-learn consortium @ Inria Foundation

EXPERIENCE

Research Engineer, Inria; Palaiseau, France — 2017 - Present

Maintenance and development of the scikit-learn library.

EDUCATION

PhD in Medical Imaging, 2012 - 2016

University of Girona, Girona, Spain

University of Burgundy, Le Creusot, France

Master in Science, 2012 - 2014

Business Innovation and Technology Management

University of Girona, Girona, Spain

Master in Science, 2009 - 2011

Computer Vision and Robotics

Heriot-Watt University, Edinburgh, Scotland

University of Girona, Girona, Spain

University of Burgundy, Le Creusot, France

SELECTED SOFTWARE CONTRIBUTIONS

Relevant information

OSSInsight: <https://ossinsight.io/analyze/glemaitre>

GitHub: <https://github.com/glemaitre>

Scikit-learn (since 2018)

Core contributor and release manager.

581 pull-requests, 848 issue comments, 4,493 reviews, 9,604 review comments

Imbalanced-learn (since 2017)

Core contributor and release manager.

277 pull-requests, 1,218 issue comments, 56 reviews, 214 review comments

SCIENTIFIC PRODUCTION

Relevant information

Google scholar: <https://scholar.google.com/citations?user=fmnnGf4AAAAJ>

4,301 citations, h-index 23

Selected publications

Traut, N., Heuer, K., Lemaître, G., Beggiato, A., Germanaud, D., Elmaleh, M., ... & Varoquaux, G. (2022). Insights from an autism imaging biomarker challenge: promises and threats to biomarker discovery. *NeuroImage*, 255, 119171.

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*, 18(1), 559-563.

Alsaih, K., Lemaitre, G., Rastgoo, M., Massich, J., Sidibé, D., & Meriaudeau, F. (2017). Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomedical engineering online*, 16, 1-12.

Lemaître, G., Rastgoo, M., Massich, J., Cheung, C. Y., Wong, T. Y., Lamoureux, E., ... & Sidibé, D. (2016). Classification of SD-OCT volumes using local binary patterns: experimental validation for DME detection. *Journal of ophthalmology*, 2016.

Lemaître, G., Martí, R., Freixenet, J., Vilanova, J. C., Walker, P. M., & Meriaudeau, F. (2015). Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Computers in biology and medicine*, 60, 8-31.

PI name	Email address			
Guillaume Lemaitre	g.lemaitre58@gmail.com			
Contractors/Consultants	Name	Year 1	Year 2	Direct Cost
Development & Maintenance	Mentees grant #1	\$ 36,000.00	\$ 36,000.00	\$ 72,000.00
Development & Maintenance	Mentees grant #2	\$ 36,000.00	\$ 36,000.00	\$ 72,000.00
			Subtotal	\$ 144,000.00
Subcontractors	Cost	Direct Cost		
Development & Maintenance	Lucy Liu - Quansight	\$ 57,960.00	\$ 62,560.00	\$ 120,520.00
Travel	Purpose	Direct Cost		
Developer sprints	Travelling fees for participants	\$ 15,000.00	\$ 15,000.00	\$ 30,000.00
Supplies & Equipment	Direct Cost			
Developer sprints	Location renting and food	\$ 15,000.00	\$ 15,000.00	\$ 30,000.00
Administrative Support	Direct Cost			
NumFOCUS	Legal & grant management	\$ 16,000.00		
Total Direct Cost				\$ 340,520.00
Indirect Costs				
Quansight Indirect Cost	15% indirect costs for Lucy Liu : - 10 % to Quansight - 5% to NumFocus			\$ 18,078.00
NumFOCUS Indirect Cost	15% indirect costs for: - Mentees' salaries - Travel - Supplies			\$ 30,600.00
Total		\$ 389,198.00		