

1. Home	2
1.1 Milestones	4
1.1.1 Sprint Plan	5
1.1.2 Sprint Retrospectives	8
1.1.2.1 Sprint 1 Retrospective	9
1.2 Specifications	10
1.2.1 Product Description	11
1.2.2 Motivational Model	13
1.2.3 User Stories	15
1.2.4 Persona	18
1.2.5 Prototype	25
1.2.6 Business Case for Cross-Team Collaboration	27
1.3 Development	29
1.3.1 Development Process	30
1.3.2 Version Control Guidelines	31
1.3.3 Test cases	32
1.3.4 Frontend workspace structure proposal	35
1.4 Resources	40
1.4.1 Contacts	41

Home

COMP90082 Hacking Materials User Interface (HA)

Product overview

In computational materials engineering, engineering new materials are accelerated by avoiding expensive and lengthy experiments to demonstrate materials' performance. Data mining methods discover better materials by searching computer-generated databases with simulations predicting using high-throughput and high-performance computing. Unfortunately, the translation of data mining methods into engineering practice is still in its infancy with some frontrunners (e.g., Tesla). A barrier is materials engineers' software tools. An engineer with materials domain knowledge needs to access and process these data efficiently to make an informed decision for eventual machine learning strategies.

The [matminer](#) python library provides a framework to simplify the process of data retrieval, feature extraction, machine learning, and visualisation. But even a python tool is a high barrier in engineering practice. Our team product aims to provide a user interface for engineering for a standard process in mat miner: Retrieving data from databases, user selection of features to be extracted within the databases, performing simple machine-learning tasks (scikit learn, Keras), and visualising results.

Goal

Provide a user interface for standard processes in MatMiner: retrieving data from a database, users selecting features to be extracted from the database, performing simple machine learning tasks (Skikit Learn, Keras), and visualizing the results.

Team Members

Name	Dr Christian Brandl	Mauro Mello Jr			
Role	Client	Supervisor			
Name	Yanan Liu (Rep)	Hongpei Lu	Jiahao Ju	Xinle Yu	Rui ZHANG
Role	Team Representative Product Developer	Product Developer	Product Developer	Product Developer	Product Developer
Student ID	1289747	1275238	1128182	1294310	1221568

Recent space activity

Rui ZHANG

Useful links

Trello	Github	Slack

Tasks

- [Edit this home page](#) - Click *Edit* in the top right of this screen to customize your Space home page
- [Create your first page](#) - Click the *Create* button in the header to get started
- [Brand your Space](#) - Click *Configure Sidebar* in the left panel to update space details and logo
- [Set permissions](#) - Click *Space Tools* in the left sidebar to update permissions and give others access



| Persona 13 •

[Frontend workspace structure proposal](#) 22 •

Yanan Liu

2 | Sprint 2 Checklist (retrieved from Mauro) 39 •

| Motivational Model 8 •



Rui ZHANG

[Home](#) 9 •

| Milestones

Sprint

[| Sprint Plan](#)

[1 | Sprint 1 Checklist](#)

[2 | Sprint 2 Checklist \(retrieved from Mauro\)](#)

Sprint Plan

Assigned to @Rui Zhang, [Xinle Yu](#), [Hongpei Lu](#)

Can download file in here: [Hacking_interface_Gantt Chart.pdf](#), see original work in [here](#).

See details on Trello board [HERE](#).

Product Preparation (Aug 1 - Aug 7)

- Team form (Aug 1-3)
- Meeting with supervisor (Aug 4)
- Create a workspace on GitHub and Trello (Aug 5)
- Create Confluence pages (Aug 6)
- Draft of the workflow (Aug 7)

Sprint 1: Inception (Aug 8 - Aug 21)

- Kick-off Meeting with Client (Aug 12)
- Motivational Model (Aug 8 - Aug 14)
- User Story & Personas Of Users (Aug 15 - Aug 21)
 - Student: Box Jelly
 - Industry: Blue Ring
 - Pro-user: Red Back
- Prototype (Aug 20 - Aug 21)
- Development Environment (Aug 15 - Aug 18)
- Development workflow (Aug 19 - Aug 21)
- Documentation (Aug 8 - Aug 21)

Sprint 2: Development (Aug 22 - Sep 19)

- Retrospective of sprint 1 (Aug 22 - Aug 24)
- Sprint planning (Aug 23 - Aug 25)
- Sprint period (Aug 26 - Sep 19)
 - US 9 View and plot the results of the model
 - Frontend ([Rui ZHANG](#), 5 days)
 - US 22 See clear annotation or explanation of data points and features
 - Frontend ([Rui ZHANG](#), 5 days)
 - US 31 Select a Machine Learning model
 - Frontend ([Rui ZHANG](#), 2 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), 3 days each)
 - Backend ([Xinle Yu Jiahao Ju](#), 5 days each)
 - US 33 Browse built-in ML models
 - Frontend ([Rui ZHANG](#), 2 day)
 - Machine learning ([Hongpei Lu Yanan Liu](#), 10 days each)
 - Backend ([Xinle Yu Jiahao Ju](#), 5 days each)
 - Testing
 - Frontend ([Rui ZHANG](#), 3 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), 3 days each)
- Task Tracking (Aug 22 - Sep 19)
- Documentation (Aug 22 - Sep 19)

Sprint 3: Development (Sep 20 - Oct 21, TBC)

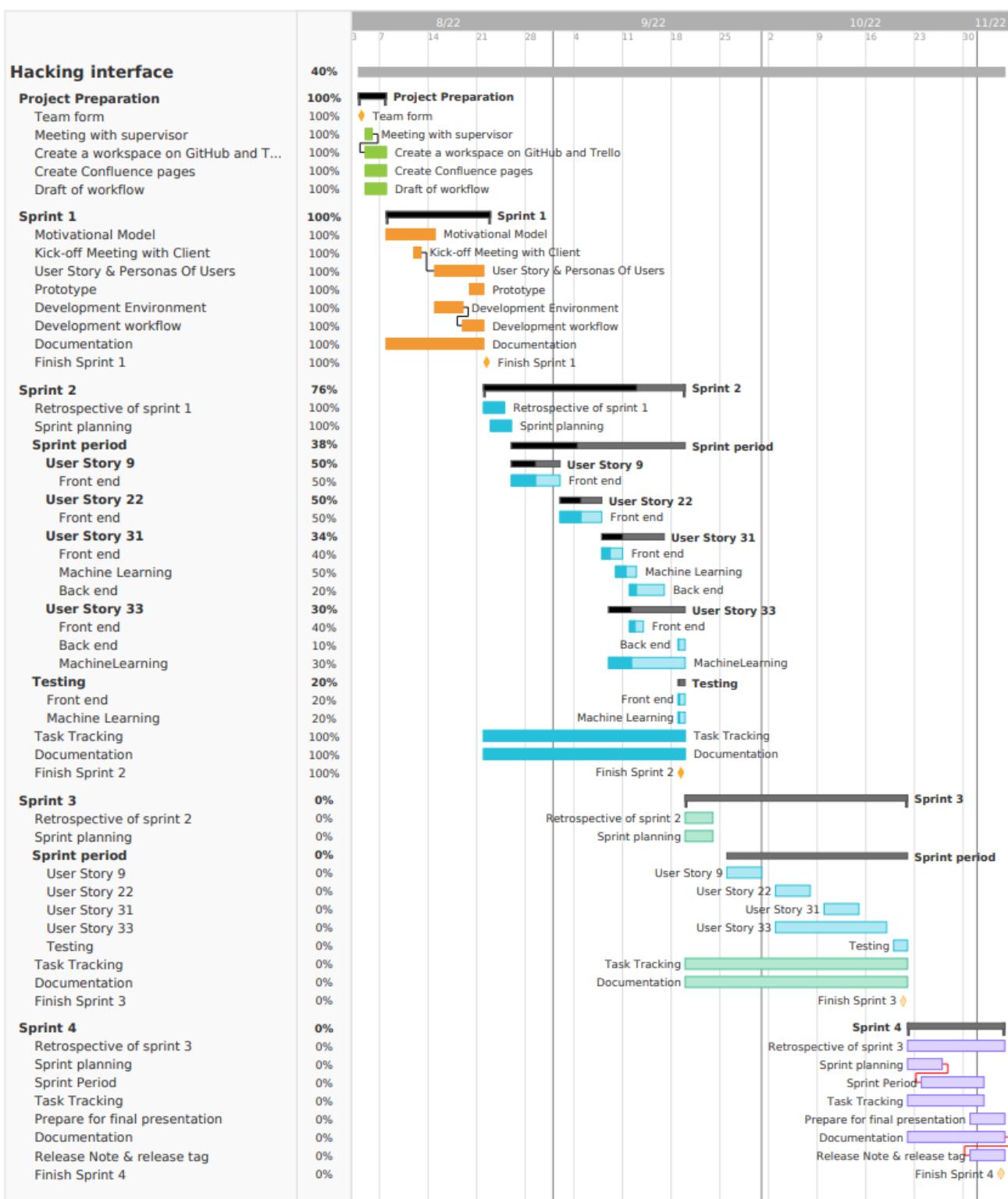
- Retrospective of sprint 2 (Sep 20 - Sep 23)
- Sprint planning (Sep 20 - Sep 23)
- Sprint period (Sep 24 - Oct 21)
 - US 8 Use different types of plotting graphs
 - Frontend ([Rui ZHANG](#), estimated 10 days)
 - US 9 View and plot the results of the model
 - Frontend ([Rui ZHANG](#), estimated 3 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), estimated 3 days each)
 - Backend ([Xinle Yu Jiahao Ju](#), estimated 5 days each)
 - US 31 Select a Machine Learning model
 - Frontend ([Rui ZHANG](#), estimated 5 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), estimated 3 days each)
 - Backend ([Xinle Yu Jiahao Ju](#), estimated 5 days each)
 - US 33 Browse built-in ML models
 - Frontend ([Rui ZHANG](#), estimated 5 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), estimated 5 days each)

- Backend ([Xinle Yu Jiahao Ju](#), estimated 5 days each)
- US 39 Select split ratio of data
 - Frontend ([Rui ZHANG](#), estimated 5 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), estimated 5 days each)
 - Backend ([Xinle Yu Jiahao Ju](#), estimated 5 days each)
- Testing
 - Frontend ([Rui ZHANG](#), estimated 5 days)
 - Machine learning ([Hongpei Lu Yanan Liu](#), estimated 5 days each)
 - Backend ([Xinle Yu Jiahao Ju](#), estimated 5 days each)
- Task Tracking (Sep 20 - Oct 21)
- Documentation (Sep 20 - Oct 21)

Sprint 4: Product (Oct 22 - Nov 4, TBC)

- Retrospective of sprint 3 (Oct 22 - Oct 26)
- Sprint planning (Oct 22 - Oct 26)
- Sprint period (Oct 24 - Nov 1)
 - Testing
- Task Tracking (Oct 22 - Nov 4)
- Prepare for final presentation (Oct 31 - Nov 4)
- Documentation (Oct 22 - Nov 4)
 - Release Note & release tag (Oct 31 - Nov 4)

Gantt Chart



Sprint Retrospectives

[Sprint 1 Retrospective](#)

Sprint 1 Retrospective

What went well	What could be done better	What we think of the client interactions
<ul style="list-style-type: none"> An important decision was made to merge the three project groups. Within the group, the division of labor is clear, the atmosphere is harmonious, and the assigned task is completed in time. The three groups learned and discussed with each other, which was of great significance for the progress of the project and the completion of Confluence. Confluence page is neat and beautiful. The communication with the Client was very harmonious, and the Client was highly satisfied with the progress of the three teams. The Supervisor plays a key role in guiding and supervising the progress of the project and the development of thinking. The use of Trello has played a positive role in the division of tasks and Sprint Plan. Shared GitHub repository and Trello board. Visibility of other team's Confluence pages. 	<ul style="list-style-type: none"> The discussion on Slack is not active enough. The process of deciding on three groups and developing the project took too much time. In the meeting of three groups, the group members were not active enough to speak. There were mistakes with the details on some Confluence pages. Our team and BoxJelly team confused about what each team should do next and what other teams are doing. No regular standup meeting. Didn't prepare a way to track everybody's work. 	<ul style="list-style-type: none"> We believe that Dr. Christian Brandl, as a Client of the project, is very cooperative in the progress of the project. He actively answered our questions and put forward his requests and ideas in the group meeting.

| Specifications

- | Product Description
- | Sprint Plan
- | Motivational Model
- | User Stories
- | Persona
- | Prototype
- | Business Case for Cross-Team Collaboration

| Product Description

Background

- In computational materials engineering, the engineering design of new materials is accelerated by avoiding costly and lengthy experiments to demonstrate the properties of materials. Data mining methods discover better materials by searching computer generated databases and making simulation predictions using high throughput and high-performance computing.
- Unfortunately, for some of the first movers, such as Tesla, translating this new thinking into engineering practice is still in its infancy. One obstacle is material engineers' software tools.
- Engineers with knowledge of the material domain need to effectively access and process this data in order to make informed decisions for the final machine learning strategy.

Goal

- Provide a user interface for standard processes in MatMiner: retrieving data from a database, users selecting features to be extracted from the database, performing simple machine learning tasks (Scikit Learn, Keras), and visualizing the results.

Sponsor

- **Dr Christian Brandl**
- Lecturer and sponsor at the University Of Melbourne.
- Completed his Ph.D. in Materials Science & Engineering at the École Polytechnique Fédérale de Lausanne EPFL and the Paul Scherrer Institute (Switzerland), followed by a post-doc in the Theoretical Division at the Los Alamos National Laboratory (USA).
- His research focuses on predictive atomistic simulation approaches to enable rational materials design.

In-scope

The product is to design and build the following:

- A web page tool to help professional user and regular user or student who has no or very little experience in machine learning analysis method to complete material analysis;

Users requirements:

- Detailed analysis of user requirements;
- Design a beautified, user-friendly interface;
- Design a login page;
- Provide hints and guidance for new users;
- Design various machine learning models in Python that allow users to select features from a database and obtain a visual result;
- Allow to add new features;
- Provide code interface for pro user;
- The workflow allows to download;
- Machine learning model accuracy reaches a sufficient level;
- Complete the improvement suggestions that may be put forward by the clients;
- At least allow 30+ users to operate concurrently;
- Report the progress and results of the product;
- Runs on Melbourne Research Cloud;
- The final product results passed the test.

Out-of-Scope

- The team has no obligation to release back-end code to the public;
- Operating maintenance is not within the scope of the team's obligations;
- The interpretability of the machine learning model does not need to be guaranteed, which means that the team members can choose the black box model as the machine learning model;
- Due to the lack of resources, the backend server cannot guarantee the stability of multiple simultaneous access.

Stakeholders

Name	Position	Internal/External	Product Role
Dr. Christian Brandl	The person who originally wanted to start this product	External	Product Client
Mauro Mello Jr	The person who directs and supervises product team members on the product	Internal	Product Supervisor

Yanan Liu	Student of the University Of Melbourne enrolled in COMP90082	Internal	Team Representative Product Developer
Hongpei Lu	Student of the University Of Melbourne enrolled in COMP90082	Internal	Product Developer
Jiahao Ju	Student of the University Of Melbourne enrolled in COMP90082	Internal	Product Developer
Xinle Yu	Student of the University Of Melbourne enrolled in COMP90082	Internal	Product Developer
Rui Zhang	Student of the University Of Melbourne enrolled in COMP90082	Internal	Product Developer

Techniques

- Web Development Techniques
 - NodeJS is a back-end JavaScript runtime environment for building scalable network applications. In this product, it is applied to build the structure of the online tool.
- Frontend
 - React is a JavaScript library for building user interfaces. It is applied to achieve designed functions.
- Backend
 - Django is a Python-based web framework. The backend process and management are based on it.
- Machine Learning
 - [MatMiner](#) is a Python library for data mining the properties of materials, the feature extraction from database will be achieved based on it.
 - [Scikit Learn](#) is a Python library for Machine Learning with various algorithms. It is used for predicting properties of materials.

Motivational Model

Assigned to [Yanan Liu](#), [Hongpei Lu](#)

Versions

Version ID	Description	Editor	Date
1.0	First version based on initial understanding of the product and first client meeting	Yanan Liu	2022-8-17
1.1	Add a goal model based on the initial do-be-feel list	Hongpei Lu	2022-8-18
2.0	All teams merged the result	Yanan Liu Hongpei Lu	2022-8-19

Version 2.0

Group cooperation work [here](#)

Do-Be-Feel-Who List

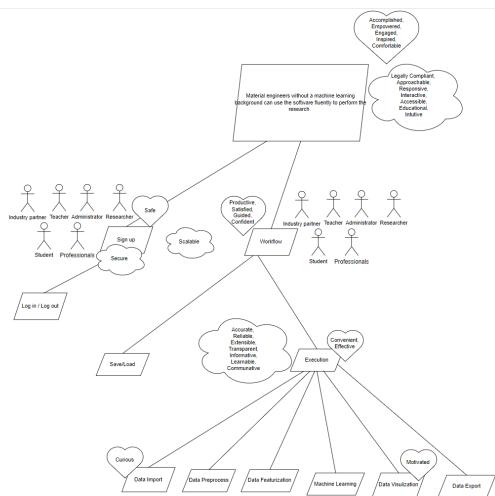
Overall goal: Material engineers without a machine learning background can use the software fluently to perform the required data analysis

Who (Roles)	Do(Functional Goal)	Be(Quality Goal)	Feel (Emotion Goal)
Students	Sign up	Accurate	Accomplished
Administrators	Log in/Log out	Approachable	Effective
Professionals	Import Data: Import data files (CSV, XES, Parquet) from local system	Reliable	Satisfied
Industry Partners	Import Data: Drag and drop import of files	Scalable	Productive
Teachers	Import Data: Create working spaces when importing	Extensible	Empowered
Researcher	Data Pre-processing: Overview of the current import data	Transparent (progress, error messages, notebook export...)	Engaged
Code maintainers	Data Pre-processing: Calculate descriptive statistics	Intuitive	Curious
	Data Pre-processing: Reduces noise and eliminates ambiguity	Accessible	Confident
	Data Pre-processing: Consider anonymized data	Secure	Comfortable
	Data Pre-processing: Standardizing data to bring it into the formatting range	Responsive	Guided
	Featurization data: Add multiple composition-based features	Learnable	Motivated
	Featurization data: Add multiple simple density features	Communicative	Convenient
	Machine Learning: Select the machine learning model to be used	Interactive	Safe
	Machine Learning: Define input data and output data: Splitting data into training, test, and validation sets	Legally Compliant	Inspired
	Machine Learning: Determining model features and training the model: Configure and adjust hyperparameters for optimum performance	Educational	
	Machine Learning: Evaluate model performance and establish benchmarks: Evaluate models using validation methods and validation datasets	Informative	
	Machine Learning: Evaluate model performance and establish benchmarks: Continuous measurement and monitoring of model performance		
	Machine Learning: Get model results: The most important features of the current ML model		
	Save / Load workflows		

Export Result: Data Visualization & Comparison - Tabular data & Plotted Graph
Export Result: Pure data table
Export jupyter notebook file
Edit python code directly in the interface
Adding more databases, machine learning methods and plot types
Export Data(input)
Maintain software

Goal model

Group cooperation work [here](#)



| User Stories

Assigned to [Rui ZHANG](#)

Versions

Version ID	Description	Editor	Date
1.0	First version based on initial understanding of the project and first client meeting	Rui ZHANG	2022-8-18
1.1	Added size estimation, priority, and justifications.	Rui ZHANG	2022-8-18
1.2	Supplement the remaining User Stories	Jiahao Ju	2022-8-18
1.3	Changed priority to Moscow priority. Added user story 5.	Rui ZHANG	2022-8-18
1.4	Delete unrelated stories. Add some general stories.	Jiahao Ju	2022-8-19
2.0	Merged user stories from all three teams	Rui ZHANG	2022-8-20
2.1	Reordered the user stories and grouped them by domain. Changed size to estimation of days	Rui ZHANG	2022-8-21
2.2	Regroup Action to new Epic and change original Epic to Domain	Yanan Liu	2022-8-21
2.3	Add epics summary table and explanation for size/priority	Rui ZHANG	2022-9-18

Version 2.3

Group cooperation work [here](#)

Note: Size was estimated at the end of sprint 1 by technically experienced group members. Size in days indicates workloads. 5-day user stories are large-sized; 3-day user stores are medium-sized; 1 day user stories are small-sized. Priority was set based on client's input.

Epic	Allocation	Total Size
Input Data	RedBack	27
Administration	BoxJelly	31
Machine Learning	BlueRing	16
Data Visualisation	BlueRing	13
Jupyter Notebook	Unallocated - lower priority	20
External Data	Unallocated - lower priority	8

ID	Role	Action	Epic	Domain		Goal	Size (days)	Priority
30	A s s a	genera l user I want to be able to view the citations for used featurizers	Input Data	Document ation	so that	I could be know more about the source of the featurizer (legally compliant)	1	1 - Must have
32	A s s a	genera l user I want to browse and select built-in featurizers	Input Data	ML	so that	I can discover ways of manipulating my data	1	1 - Must have
34	A s s a	genera l user I want to browse built-in datasets	Input Data	Server	so that	I can discover data to experiment with	1	1 - Must have
19	A s s a	student I want to quickly browse the Materials available in the database for retrieval and simulations	Input Data	UI	so that	I can quickly perform queries.	3	1 - Must have
21	A s s a	genera l user I want to be able to select datasets from existing databases	Input Data	Data Manipulati on	so that	I do not have to worry about how the data is loaded	3	1 - Must have
37	A s s a	genera l user I want to be able to preview the input data	Input Data	UI	so that	I could explore the data	1	2 - Should have
25	A s s a	genera l user I want to Select specific features from a dataset	Input Data	Data Manipulati on	so that	I can improve the precision of my model	3	2 - Should have

13	A s a	Pro user	I want to	add new features	Input Data	ML	so that	they can be reused in the future	5	2 - Should have
28	A s a	genera l user	I want to	be able to reference / view citation for original data sources	Input Data	Document ation	so that	I can retrieve data.	1	3 - Could have
18	A s a	pro user	I want to	be able to apply new featurizers	Input Data	ML	so that	I can create new features	3	3 - Could have
1	A s a	student	I want to	clean and tune data input	Input Data	Data Manipulati on	so that	I have less noise on visualizations.	5	3 - Could have
29	A s a	student	I want to	save project specific data/checkpoints	Administra tion	Server	so that	I can pick up where I left off for specific projects	1	1 - Must have
35	A s a	pro user	I want to	export model selections, parameters, and data flows	Administra tion	Server	so that	I can save my work and share it with others	1	1 - Must have
36	A s a	pro user	I want to	import exported model selections, parameters, and data flows	Administra tion	Server	so that	I can continue work I had previously saved	1	1 - Must have
20	A s a	student	I want to	Create an account using single-sign on, restricted to the *.unimelb.edu.au domain	Administra tion	Server	so that	my research remains private	3	1 - Must have
23	A s a	pro user	I want to	Control job execution	Administra tion	Server	so that	I can start, view progress of, and cancel jobs related to my project	3	1 - Must have
10	A s a	pro user	I want to	be able to opt in to pro-user features	Administra tion	UI	so that	I can access pro user features	5	1 - Must have
38	A s a	pro user	I want to	have my pro user settings persist on each visit	Administra tion	Server	so that	I don't have to reconfigure settings to use the features I need	1	2 - Should have
24	A s a	student	I want to	receive provided hints and guidance for new users	Administra tion	UI	so that	I can quickly learn how to use software	3	2 - Should have
14	A s a	pro user	I want to	easily find and read documentation on the pro features	Administra tion	Document ation	so that	I can use them with ease	5	2 - Should have
17	A s a	pro user	I want to	Be kept informed about job status	Administra tion	Server	so that	I can avoid polling my workspace to check for results	3	3 - Could have
6	A s a	pro user	I want to	have access to more processing power	Administra tion	Server	so that	I can run more complex operations or use more data	5	3 - Could have
31	A s a	genera l user	I want to	able to select a Machine Learning model	Machine Learning	ML	so that	I could use it to train and run the data	1	1 - Must have
33	A s a	genera l user	I want to	browse built-in ML models	Machine Learning	UI	so that	I can discover ways of manipulating my data	1	1 - Must have
39	A s a	user	I want to	be able to select split ratio of data	Machine Learning	Data Manipulati on	so that	to train and test the model	1	2 - Should have
26	A s a	pro user	I want to	have the option to change the hyperparameters used in the machine learning model	Machine Learning	ML	so that	I can fine tune my test results.	3	2 - Should have
15	A s a	pro user	I want to	be able use additional ML models	Machine Learning	ML	so that	I can improve accuracy	5	2 - Should have
7	A s a	pro user	I want to	combine multiple models together	Machine Learning	ML	so that	I can model more complex data manipulations	5	3 - Could have
22	A s a	genera l user	I want to	see clear annotation or explanation of data points and features	Data Visualisati on	UI	so that	I can understand the results of the analysis	3	1 - Must have
8	A s a	student	I want to	use different type of plotting graphs	Data Visualisati on	UI	so that	I have flexibility to visualize data according to my needs.	5	1 - Must have
9	A s a	genera l user	I want to	able to view and plot the results of the model	Data Visualisati on	UI	so that	I could analysis and visualise the effects of the model	5	1 - Must have
12	A s a	student	I want to	export my work to a Jupyter Notebook	Jupyter Notebook	Server	so that	I can extend my work beyond the capability of the application	5	2 - Should have
2	A s a	genera l user	I want to	attach comments to workflow objects	Jupyter Notebook	UI	so that	I can document my work	5	3 - Could have

4	A s a	Pro user	I want to	edit python code on the interface	Jupyter Notebook	UI	so that	I can have control how the ML algorithms works	5	3 - Could have
5	A s a	Pro user	I want to	upload my own script (in python) if possible	Jupyter Notebook	Server	so that	I can extend the tool to support custom models and featurizers	5	3 - Could have
27	A s a	pro user	I want to	be able to access new databases	External Data	Data Manipulati on	so that	I can access additional data	3	2 - Should have
3	A s a	Pro user	I want to	be able to add new datasets in the future	External Data	Data Manipulati on	so that	if there's a new dataset that can be used on a new project, it can be added instantly	5	3 - Could have
11	A s a	student	I want to	analyze the relationship between different features		ML	so that	I can identify which features I need to select for my analysis	5	2 - Should have
16	A s a	genera l user	I want to	add specific materials to the workflow for analysis		Data Manipulati on	so that	compare the performance of the specific material my client or I choose with other material	3	3 - Could have

| Persona

Assigned to [Xinle Yu](#)

Version	Description	Editor	Date
1.0	First version based on initial understanding of the project and first client meeting	Xinle Yu	2022-8-19
1.1	Round off the sentences and adjust some description	Xinle Yu	2022-8-21
1.2	Combine three team work together	Yanan Liu	2022-8-22
2.0	Adjust technical skills and edit the text for fitting the change of skills	Xinle Yu	2022-9-18
2.1	Combine three team work together	Yanan Liu	2022-9-19

Persona 1: Student (By: BoxJelly)

Assol Anahita

age: 22

residence: Melbourne

education: Material Engineering

occupation: Student

marital status: Single



"It's SO time consuming to do material researches and get decent results through just a semester."

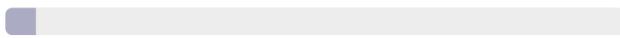
Motivation: As a material engineering graduate student, Assol gets frustrated and demotivated when she can't make sense of the data she has because she doesn't have a tool or sufficient programming/machine learning skills to process the material data. She is also frustrated that she can't use machine learning algorithms to help her engineer new materials even though she is told by her supervisor that this idea works in theory.

Comfort With Technology

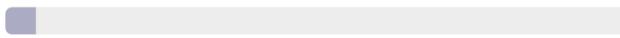
PROGRAMMING WITH PYTHON



MACHINE LEARNING



CLOUD BASED STORAGE



MATERIAL SCIENCE



Criteria For Success:

Assol can perform materials data requests/retrievals and accurate materials property predictions supported by Machine Learning technology with easy to follow steps button clicks user interface.

Needs

- Easy-to-use interface Material science data processing and retrieval application
- A tool to predict property of a material with assistance of Machine Learning technology without prior knowledge of Python and Machine Learning programming

Wants

- A data mining application that helps her researches
- A better understanding on how Machine Learning can help her to learn more about a material
- Ability to use ML algorithms as a black box
- Freedom to select features on her own terms
- A tool to accelerate research progress

Values

- Convenience
- Quickness
- Safety
- Understandable

Fears

- Spends hours without getting anything done because she neither has an adequate tool to do data mining, nor the programming skill to analyse the data herself
- Have to conduct countless experiments to figure out the properties of the materials

- Hard to choose suitable ML algorithms

Persona 2: Pro User (By: RedBack)

Alex

age: 45

residence: Melbourne

education: Masters Degree in Physics

occupation: Materials Engineer

marital status: Divorced without kids



"There has got to be a better way to do this."

Motivation : As an experienced Materials Engineer, Alex's job requires him to narrow down candidate materials by performing physical experiments to choose a material which can takes years to do. He needs a tool that can speed up the process by narrowing down candidate materials for experimentation using Machine Learning and simulations.

Comfort With Technology

PROGRAMMING WITH PYTHON



MACHINE LEARNING



CLOUD BASED STORAGE



MATERIAL SCIENCE



Criteria For Success:

Alex can find the right materials efficiently, with accurate results and that matches the client's requirements.

Needs

- Products to accelerate his workflow
- Access to wide variety of related tools and resources

Wants

- Suitable models and featurizers for different use cases
- Demonstrate reproducible results to his clients
- Share resources with others
- Refining generated workflow to reuse

Values

- Extensibility
- Accuracy
- Reliability
- Responsiveness
- Scalability
- Transparency

Fears

- Tool is too inflexible
- Losing access to progress on his work
- Not being able to verify his results
- Not having support with the tool

Persona 3: Industry User (By: BlueRing)

Gray Zhou

age: 28

residence: Ningde, Fujian, China

education: Master of Material Engineering

occupation: R & D Engineer of Polymer

marital status: Single

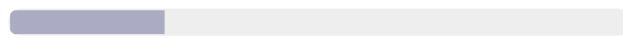


"It is fantastic to apply a multi-function online tool with ML methods if it is efficient and reliable. Nobody will refuse a tool that can save his time"

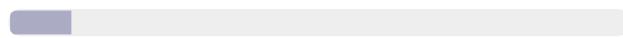
Motivation: Gray Zhou is a R & D Engineer of polymer in a battery factory. His work is searching for better materials for battery production. Gray spends lots of time testing different materials, but some of tests are waste of time because of the poor performance observed. He needs a system that can predict some useful properties of materials so that he can remove samples with low predicted performance and boost the research. His company provides some solutions, but they are awkward and only have limited functions.

Comfort With Technology

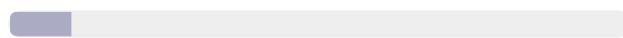
PROGRAMMING WITH PYTHON



MACHINE LEARNING



CLOUD BASED STORAGE



MATERIAL SCIENCE



Criteria For Success:

Provide a website or online-tool with quick, visual interface which can help him in daily development of new materials.

A successful product should help him save noticeable time on data processing and provide reliable prediction of properties.

Needs

- Retrieve and extract required data, process the data with ML methods to get some properties
- Provide graphs which can be modified with interface about predicted properties
- Help finding the material with best predicted properties

Wants

- Ability to interact with the graph to further compare several materials in detail
- Upload data from his lab for predicting
- Explain what ML method the system applied and how it helps the prediction
- Continue his work on mobile devices without gaps of interaction

Values

- Easy to get started on both desktop and mobile
- Efficient back-end process
- Abilities to select functions and filter results
- Well organized visualization of interface and

Fears

- Not enough guidance in the web or tool so him may feel confused to find functions he wants.
- Lacking understanding of what the system done, then reducing the confidence level of his report
- Frequently unavailability

grids

▼ Frequency and Priority

| Prototype

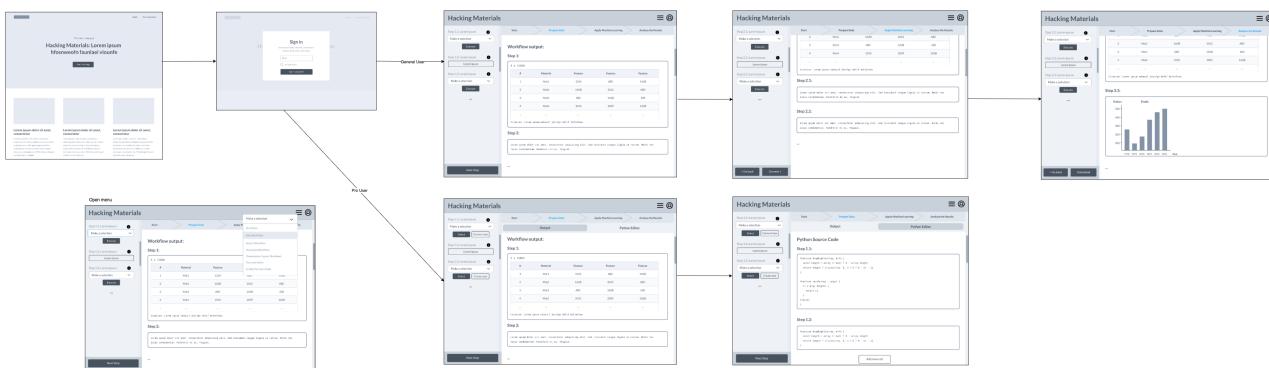
Assigned to [Yanan Liu](#)

Versions

Version ID	Description	Editor	Date
1.0	First version based on initial understanding of the project and client meeting Made in collaboration with team redback and boxjelly	Yanan Liu Ghina Yashar Felipe Leefu Huang Lin	2022-8-20

Group cooperation work, work in [marvel](#)

Version 1.0



Descriptive Notes recorded by team redback

Descriptive Notes

- Landing page:
 - Static page with information about the app and project
 - Link to access the app
 - On click, it opens a login modal
 - Once the user is logged in, they're redirected to the app
- Single page app:
 - Top bar:
 - User profile button at the top opens a menu to give the user the option to log out
 - Menu button at the top has options to import or save a workflow, download it in different formats, start over, a link to the documentation and a toggle to enable pro view.
 - General user:
 - The workflow is divided into major and minor steps. Each major step would have its own page. The user can go back and forth between the major steps as needed.
 - Left panel:
 - All the minor steps are numbers and named to guide the user
 - Inputs can be of different types
 - Each step has a tooltip button that would open a modal with guidance information about the step
 - The steps and options in the left panel should always be the same no matter what selections the user made in previous steps. Any step that requires customised inputs would open in a modal.
 - Example 1: Step 3.1 might be "Selecting a plot type". As there is a known, limited list of different plot types, this step may be a drop-down menu that is displayed directly in the left panel.
 - Example 2: Step 3.2 might be customising the selected plot's configuration options. As different plot types may need different configuration options, these options will not be displayed in the panel directly. Instead, the panel will include only a button that says "Configure plot", which would open a modal with the specific options applicable to the selected plot type.

- Pinned buttons at the bottom of the panel: navigate between the different major steps. Last step page may also have a button to download the full workflow.
- Viewing window:
 - At the top of the viewing window, the user can see the progression of major steps with the current step highlighted.
 - The output of each minor step is labelled with the step number and contained inside a box. The output inside the box is the same output produced by running the python code, simply copied over for transparency.
 - The outputs from the previous pages are also always displayed, so it's not just the outputs of the current page.
 - Where a resource with citations is used, the citations will be automatically printed after the output of the step where the resource was selected.
- Pro user:
 - Left panel: has all the same options as a general user, plus additional buttons to configure their own settings as needed
 - Viewing window: the window has 2 tabs:
 - Output: same as the viewing window of the general user
 - Python source code:
 - An editable view of all the code generated by their selections, looks similar to a Jupyter notebook.
 - User can add new cells as desired
 - Brings up the following question: what happens if the user edits the code generated by one of the steps? This may lead to inconsistencies between what is shown in the step's input field and what the code now actually does. This is an implementation decision so is not a major concern right now, but one option that we decided to show in the prototype is that the step's input in the left panel would change to say "Custom" or something similar, indicating that the configuration was changed.

| Business Case for Cross-Team Collaboration

Three teams cooperation work original in [here](#), and formatted document can be downloaded here [HA Software Project - Cross-team collaboration - Business case.pdf](#)

COMP90082 Software Project SM2 2022

Project Hacking Materials Cross-Team Collaboration Model Proposal

The current situation

We currently have three teams working with the same client, Dr. Christian Brandl, on the project Hacking Materials ("HA"). The client expressed that he would not be interested in three different versions of the same product, and would prefer we work together to be able to build one more complete final product.

Additionally, it is clear from our discussions with the client that any resulting single product will be difficult to decompose into totally independent components. This means that the deliverables produced by each team will be dependent on those produced by other teams, requiring collaboration on design, development, and project management tasks.

This issue is compounded by the fact that each team has a separate workspace set up by the university, in particular the separate GitHub repositories. This makes it extremely difficult to share resources including source code and development resources and to determine how much progress is being made by each team, and on which tasks.

Our biggest concern is that our teams will be assessed separately. Working on the same product makes us inherently reliant on each other, which could affect the assessment of each team. We understand that this complicates the situation, and have come up with a working agreement that we believe would allow us to deliver the product that the client wants while minimizing the risks to the assessable components of the project.

Obstacles

We have identified several issues with the way the project is currently being approached, including the following:

- The project scope is larger than what could realistically be accomplished by one student team.
- If the teams were to work completely separately on different parts of the product, it will be difficult to organize ourselves in a way that ensures the needs of all teams are met.
- Finding common availability between 15 people for meetings is difficult. As such, cross-team communication so far has been done mainly through team representatives. This approach is slow, high effort for team representatives (as they need to be aware of everything their teams are doing at all times) and creates an increased risk of information getting lost in transmission.

Proposed work structure/way of working

The structure we are proposing is as follows:

- The teams would work on the same GitHub repository.
- Each team works as a full-stack team focusing on a particular epic of stories,
- Someone from each team is involved in each technical area of the product (Backend/ Frontend/ Machine Learning).
- The members of the different teams who are within the same technical area cooperate to ensure they follow the same standards and processes. This will help spread some of the cross-team communication load away from the team representatives.
- Each team will need to review the other teams' work to ensure that it doesn't interfere with or affect their own work.
- Ownership of epics will be allocated to teams, but the teams will work together to ensure all high-priority user stories are delivered first. This creates fewer dependencies between teams and allows each team to deliver fully functional components without relying on the others.
- Each team will be responsible for its own Confluence space, but the teams will maintain a similar structure to make navigation easy. Major structural changes will be reported and shared with the other teams during cross-team meetings.
- The teams will share certain deliverables within their Confluence spaces so that the client would not have three different versions of the same document.
- Sprint documents exported from each team's Confluence space will be included within their own folder in the shared GitHub repository.
- The teams will take turns organizing meetings with the client, supervisor, and each other.

Requested actions

- **The decision from teaching staff:** University administration hasn't made a decision on assessment criteria, or whether the teams can collaborate on this project. This prevents us from adopting a shared working model because we are not sure whether our project submissions will be accepted by the university, or whether we will be otherwise penalized for taking this approach. This was discussed with our supervisor, Mauro, and it was agreed that a response would need to be available by Tuesday 3 pm.
- **Shared GitHub repository:** If our proposal is accepted, we would need a shared GitHub repository so that Dr. Christian Brandl will not have three copies of each deliverable.

| Development

Development Process

Proposed by Red Back team and approved among three teams

- All tests are required to pass in CI before landing a pull request
- Sprint lifecycle:
 - Sprint Kickoff:
 - Review and re-estimate tasks: user stories get t-shirt size and priority
 - Development:
 - Feature kickoff:
 - Specify test cases and acceptance criteria
 - Tasks are estimated in the number of days to complete using the magic estimation approach
 - Code reviews:
 - Require tests to pass in CI before merging
 - At least one other RedBack member must approve the pull request before it can be merged
 - At least one BoxJelly member working on the same piece of technology must approve the pull request before it can be merged
 - At least one BlueRing member working on the same piece of technology must approve the pull request before it can be merged
 - All test cases and acceptance criteria identified in kickoff must be satisfied
 - Use auto-formatters to maintain code quality
 - Branching
 - Use the format `feature/t-<ticket>` as a feature branch template, where `<ticket>` is the Trello card number
 - `<username>/idea` for scratch/experimenting branches
 - `main` is the main branch
 - We will follow the following guidelines: <https://www.atlassian.com/git/tutorials/comparing-workflows/gitflow-workflow> (we probably only need main, release, scratch, and feature branches)
 - Release management:
 - Deploy as required

Version Control Guidelines

Group cooperation work

Code quality guideline

- Code should be able to run, there's no syntax error
- If test(s) are included in the pull request, they should pass
- All previous tests should still pass with the new changes in the pull request
- A branch should not be put up for a pull request if it has merge conflicts with the main, all conflicts should be resolved before that
- Code should be understandable and contain documentation
- The code should not include global state

Code review guideline

- Has to be reviewed by 1 member from the other 2 teams
- Pull request should have assigned reviewers within 24h once it's no longer DRAFT and should be reviewed (either approve or require changes) within 48h
- Pull requests should include committing messages describing the work you've done and steps you did to verify your work
- Pull request will be reviewed by using the Review Changes button under the Files changed tab
- Pull requests should only be merged/rebased by the creator of the pull request
- Branch that has been successfully merged/rebased to main should be deleted by the creator of the pull request
- Reviewers should follow the work quality guideline to review the code

Acceptance criteria definition guideline

- It should be defined from the user's point of view
- It should contain a list of steps to test the desired functionality

Definition of done (for user story)

- Acceptance criteria should be defined for the user story and pass
- All related code has passed code review and merged to the main

Test cases

Versions

Version ID	Description	Editor	Date
1.0	Add test cases for the US 31	Hongpei Lu	2022-9-15

US 31: As a general user, I want to be able to select a Machine Learning model.

TC 01: Select a model (successful)

Test Type: Functional	Execution Type: Manual
Objective: Verify if a machine learning model is correctly selected.	
Setup: The IO stream has been established between the front end and the back end.	
Pre-Condition: 1. The user has inputted the data or selected the data. 2. The user has selected the target feature.	
Notes: [1] Select the machine learning model in the interface. Must select a choice in the given place. Do not select the models that are out of scope in the given choices (Linear regression model and Random forest model). [2] Submit the information. * Application starts machine learning using given input dataset and selected feature.	
Time constraint: Minimum: 10 min Maximum: 60 min	

TC 02: Select a model (unsuccessful)

Test Type: Functional	Execution Type: Manual
Objective: Verify if a machine learning model is correctly selected.	
Setup: The IO stream has been established between the front end and the back end.	
Pre-Condition: 1. The user has inputted the data or selected the data. 2. The user has selected the target feature.	

Notes:

- [1] Try not to select the machine learning model in the interface.
 [2] Submit the information.
 * Application cannot learn without a target selected machine learning model.

Time constraint:

Minimum: 10 min
 Maximum: 60 min

TC 03: Enter a split ratio of the dataset (successful)

Test Type: Functional	Execution Type: Manual
Objective:	
Verify if a machine learning model can split a dataset as required.	
Setup:	
The IO stream has been established between the front end and back end.	
Pre-Condition:	
<ol style="list-style-type: none"> 1. The user has inputted the data or selected the data. 2. The user has selected the target feature. 3. The target machine learning method has been selected. 	
Notes:	
<p>[1] Fill in the split ratio in the given place. * Split ratio cannot be special chars or letters. * Split ratio must be a number that is bigger or equal to 0 and less or equal to 1. [2] Submit the information. * Application starts machine learning using given input dataset and selected feature.</p>	
Time constraint:	
<p>Minimum: 10 min Maximum: 60 min</p>	

TC 04: Enter a split ratio of the dataset (unsuccessful)

Test Type: Functional	Execution Type: Manual
Objective:	
Verify if a machine learning model can split a dataset as required.	
Setup:	
The IO stream has been established between the front end and the back end.	
Pre-Condition:	
<ol style="list-style-type: none"> 1. The user has inputted the data or selected the data. 2. The user has selected the target feature. 3. The target machine learning method has been selected. 	

Notes:

- [1] Try not to fill in the blank.
- [2] Try to fill in the blank using an invalid content-type:
 - [2.1] Try to fill in the blank using special chars.
 - [2.2] Try to fill in the blank using letters.
 - [2.3] Try to fill in the blank using a number that is bigger than 1.
 - [2.4] Try to fill in the blank using a number that less than 0.
- [3] Submit the information.

*Split ratio cannot be special chars or letters.

*Split ratio must be a number that is bigger or equal to 0 and less or equal to 1.

Time constraint:

Minimum: 10 min

Maximum: 60 min

Frontend workspace structure proposal

Proposed by team Red Back and approved among three teams

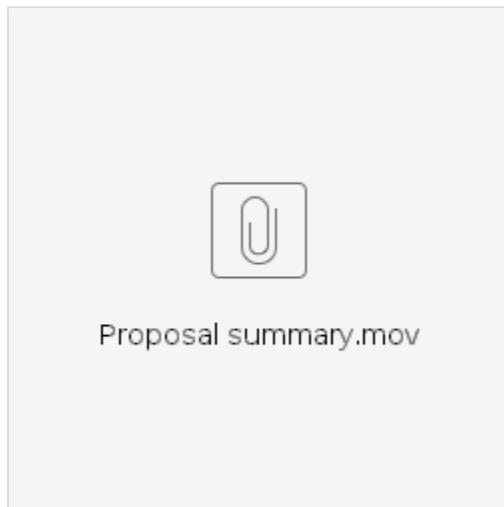
Proposed by	Ghina Yashar
Presented to	All frontend contributors from teams RedBack, BlueRing and BoxJelly
Proposal date	September 1 2022
Status	APPROVED - September 2 2022
Approvers	<ul style="list-style-type: none">■ Mamta Lopes(RedBack): 1/9/2022■ Felipe Lin (BoxJelly): 2/9/2022■ Rui Zhang (BlueRing): 2/9/2022

Proposed Structure

Please note: All the names used below can be replaced if needed, the focus of this proposal is more on the structure rather than the naming.

Summary video

If you don't like reading, please watch the video below for a quick overview of the proposed structure. The sample code snippets shown in the video are copied below as well.



Summary in writing

The structure I'm proposing would follow this rough directory tree:

Sample directory structure

```

|_assets
|_src
  |_components
    |_exampleComponent
      |_examples.tsx
      |_index.tsx
      |_test.tsx
      |_styled.tsx
    |_dropdownSelectStepType
      |_examples.tsx
      |_index.tsx
      |_test.tsx
      |_styled.tsx
    ...
  |_steps
    |_datasetSelection // (e.g.)
      |_index.tsx
      |_test.tsx
      |_HelpModal
        |_index.tsx
        |_styled.tsx
      ...
    ...
  ...
  |_sections
    |_appHeader
      |_index.tsx
    ...
    |_appBody
      |_InputPanel
        |_index.tsx
      ...
      |_ViewingWindow
        |_index.tsx
      ...
    |_index.tsx
  |_appFooter // amendment suggested by Felipe
  ...
  |_App.tsx
  ...
|_package.json
|_README.md
...

```

The main ideas of this are as follows:

Sections

By referring to the [low-fidelity prototype](#) created earlier in the project, we divide the main application page into 2 main sections:

- Header: the top bar, which does not need to have context of what stage the user is up to and what's happening at any given point.
- Body: Includes 2 subsections that both need to know which stage the user is at (e.g. "Pre-process data" or "Apply machine learning"):
 - Left-side panel: named in the structure as `InputPanel`. Example code for this panel and how it shows the workflow steps is included in the sample code section below.
 - Main window on the right: named in the structure as `ViewingWindow`

AMENDMENT - 2 SEPTEMBER 2022

Felipe Lin (HA-BoxJelly) suggests potentially adding an `appFooter` as well. No objections to this so far.

Hacking Materials **SECTION: HEADER** ≡ ⌂

Step 2.1: Lorem ipsum i

Make a selection

Execute

Step 2.2: Lorem ipsum i

Lorem ipsum

Step 2.3: Lorem ipsum i

Make a selection

Execute

...

SECTION: BODY

Step 2.1:

Lore ipsum dolor sit amet, consectetur adipiscing elit. Sed tincidunt congue ligula in rutrum. Morbi nec lacus condimentum, hendrerit mi eu, feugiat.

Step 2.2:

Lore ipsum dolor sit amet, consectetur adipiscing elit. Sed tincidunt congue ligula in rutrum. Morbi nec lacus condimentum, hendrerit mi eu, feugiat.

...

< Go back Go

Hacking Materials ≡ ⌂

Step 2.1: Lorem ipsum i

Make a selection

Execute

Step 2.2: Lorem ipsum i

Lorem ipsum

Step 2.3: Lorem ipsum i

Make a selection

Execute

SUB-SECTION: INPUT PANEL ...

SUB-SECTION: VIEWING WINDOW

Lore ipsum dolor sit amet, consectetur adipiscing elit. Sed tincidunt congue ligula in rutrum. Morbi nec lacus condimentum, hendrerit mi eu, feugiat.

Lore ipsum dolor sit amet, consectetur adipiscing elit. Sed tincidunt congue ligula in rutrum. Morbi nec lacus condimentum, hendrerit mi eu, feugiat.

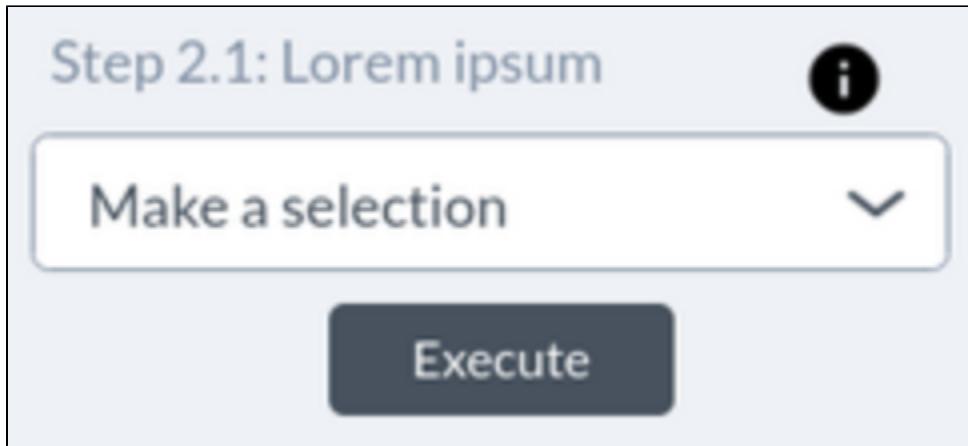
...

< Go back Go

Components

This folder will mainly contain all reusable components, e.g. Button, Tooltip, Modal, etc.

Notably, some of these reusable components would be "step types", e.g. `DropdownSelectStepType`. This example step type refers to the entire object shown below, including a step number, title, tooltip, dropdown list, button and whatever else may be needed. We would create this as a reusable component because many steps have similar requirements, e.g. selecting a dataset and selecting a featurizer should both be dropdown list type steps.



Steps

The word "steps" in this section refers specifically to the workflow steps that would be shown in the input panel, e.g. Dataset Selection step, Featurizer Selection step, etc.

A separate folder is created for these so that there would be a clear pattern that is easy to follow whenever more steps need to be added. Each step would use a step type component that is imported from the `/components` folder. E.g. the `DatasetSelectionStep` would use the `DropdownSelectStepType`, as shown in the sample code snippet below.

Sample code snippets

Sample src/steps/datasetSelection/index.tsx

```
import DropdownSelectStepType from '../../../../../components/dropdownSelectStepType';
import HelpModal from './HelpModal';

...
const DatasetSelectionStep = (props) => {
    ...
    const STEP_KEY = "dataset_selection"

    const options = api_call_here() // calls backend API to get the dataset options

    const onSubmit = selected_value => send_to_backend() // send to backend using api

    return (
        <DropdownSelectStepType
            stepNumber={props.stepNumber}
            title="Select Dataset"
            description="bla bla"
            tooltipContent={HelpModal}
            options={options}
            onSubmit={onSubmit}
        />
    );
};
```

Sample src/sections/appBody/InputPanel/index.tsx

```
import DatasetSelectionStep from '../../../../../steps/datasetSelectionStep';
import FeatuirzerSelectionStep from '../../../../../steps/featuirzerSelectionStep';
...

const InputPanel = (props) => {
    ...
    const { stage } = props;

    if (stage === 1) {
        return (
            <div>
                <DatasetSelectionStep
                    stepNumber="1.1"
                    data={data}
                    handleChange={handleChange}>
                />
                <FeatuirzerSelectionStep
                    stepNumber="1.1"
                    data={data}
                    handleChange={handleChange}>
                />
                ...
            </div>
        );
    } else if (stage === 2) {
        return (
            <div>
                ...
            </div>
        );
    }
};
```

| Resources

GitHub: <https://github.com/COMP90082-2022-SM2/HA-BlueRing>

Trello: <https://trello.com/b/Q1AZdIXb/sprint-1>

Matminer: https://nbviewer.org/github/hackingmaterials/matminer_examples/blob/main/matminer_examples/index.ipynb

LMS Project: https://canvas.lms.unimelb.edu.au/courses/126944/pages/hacking-materials-user-interface-code-ha?module_item_id=3996633

Dr. Christian Brandl information: <https://findanexpert.unimelb.edu.au/profile/826808-christian-brandl>

Motivational Model: <https://momo-staging.eresearch.unimelb.edu.au/dashboard>

| Contacts

Teammates

Name	Email address	Student ID
Yanan Liu (Team rep)	yananli7@student.unimelb.edu.au	1289747
Hongpei Lu	hongpeil@student.unimelb.edu.au	1275238
Jiahao Ju	jiahaoj1@student.unimelb.edu.au	1128182
Xinle Yu	xinley@student.unimelb.edu.au	1221568
Rui Zhang	rzzhan2@student.unimelb.edu.au	1294310

Client

Name	Email address
Dr. Christian Brandl	christian.brandl@unimelb.edu.au

Teaching Staff

Name	Email address
Mauro Mello Jr (supervisor)	mauro.mellojr@unimelb.edu.au
Eduardo Oliveira	eduardo.oliveira@unimelb.edu.au

Team RedBack

Name	Email address
Alastair Daivis (Team rep)	adaivis@student.unimelb.edu.au
Sanjeevani Avasthi	savasthi@student.unimelb.edu.au
Mamta Lopes	mlopes@student.unimelb.edu.au
Ghina Yashar	gyashar@student.unimelb.edu.au
Chunbaixue Yang	chunbaixue@student.unimelb.edu.au
Confluence Page: Home	

Team BoxJelly

Name	Email address
Zhaoqi Wang (Team rep)	zhaoqiw@student.unimelb.edu.au
Dara O hEidhin	doheidhin@student.unimelb.edu.au
Yaoming Xuan	yaomingx@student.unimelb.edu.au
Radhimas Djan	djanr@student.unimelb.edu.au
Felipe Leefu Huang Li	fleefuhuangl@student.unimelb.edu.au
Confluence Page: Home	