1. Home 2
1.1 | Specifications 4
1.1.1 | Project Overview 5
1.1.2 | Personas 6
1.1.3 | Motivational Model 10
1.1.3.1 Do/Be/Feel lists 11
1.1.3.2 Goal Model 13
1.1.4 | User Stories 15
1.1.5 | Low-fidelity Prototype 23
1.1.6 | Notes on provided material 24
1.2 | Development 25
1.2.1 | Development Process 26
1.3 | Assessments 27
1.3.1 | Cross-Team Collaboration Business Case 28
1.4 | Resources 29
1.4.1 | Links and Tools 30
1.4.2 | Contact Information 31

Home

Project Overview

Materials science is a complex field that focuses on the study of physical materials and their properties. MatMiner is an existing Python library that combines materials data with Machine Learning strategies and models to study material properties without the need for time-consuming physical experimentation. Our project, Hacking Materials ("HA"), aims to build an easy and intuitive user interface to the MatMiner library to eliminate the need for its users to have substantial Python or machine learning knowledge or experience.

Materials engineering is a field in which physical materials (e.g. metals, ceramics, polymers, composites, etc.) are studied to understand their composition, characteristics and properties. In many industries, such as mining, manufacturing and others, finding the right material for each job is essential for success. Materials engineers and scientists follow many different methods to compare candidate materials and make recommendations based on how they each satisfy the specific use case requirements. Some of these methods include physical experimentation, which can in some cases take up to 20 years to complete.

To avoid this, computer simulations (based on existing databases of known material properties and machine learning algorithms) can be used to compare materials in a much more efficient way. One such tool for this approach is the Python library MatMiner, which allows easy access to ready-made datasets and integrates well with other machine learning Python libraries. However, to use this library, the user must have a substantial level of specialised knowledge in machine learning and programming, which most materials engineering do not have. To help them overcome this, the solution proposed by the client describes a simple and intuitive user interface that would act as a bridge between the user and the MatMiner library in the backend. The vision of this product is to make machine learning methodologies more accessible within the materials science and engineering industry as a whole, minimising the time and financial costs involved and leading to a more efficient industry.

Project Personnel

	
Dr Christian Brandl	**Mauro Mello Jr**
Client	Supervisor

Useful Links

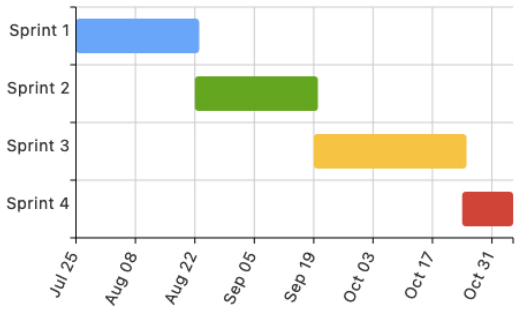
Trello	GitHub

Key Pages

| Specifications	**| Meeting**
| Assessments	**| Resourc**

			
Alastair Daivis (Alastair Daivis) Team Representative	**Sanjeevani "Sanjee" Avasthi** (Sanjeevani Avasthi)	**Mamta Lopes** (Mamta Ritesh Lopes)	**Ghina Yas** (Ghina Yas

Project Timeline



| Specifications

- | Project Overview
- | Personas
- | Motivational Model
- | User Stories
- | Low-fidelity Prototype
- | Notes on provided material

| Project Overview

The Problem

Materials engineering is a field in which physical materials (e.g. metals, ceramics, polymers, composites, etc.) are studied to understand their composition, characteristics and properties. In many industries, such as mining, manufacturing and others, finding the right material for each job is essential for success. Materials engineers and scientists follow many different methods to compare candidate materials and make recommendations based on how they each satisfy the specific use case requirements. Some of these methods include physical experimentation, which can in some cases take up to 20 years to complete.

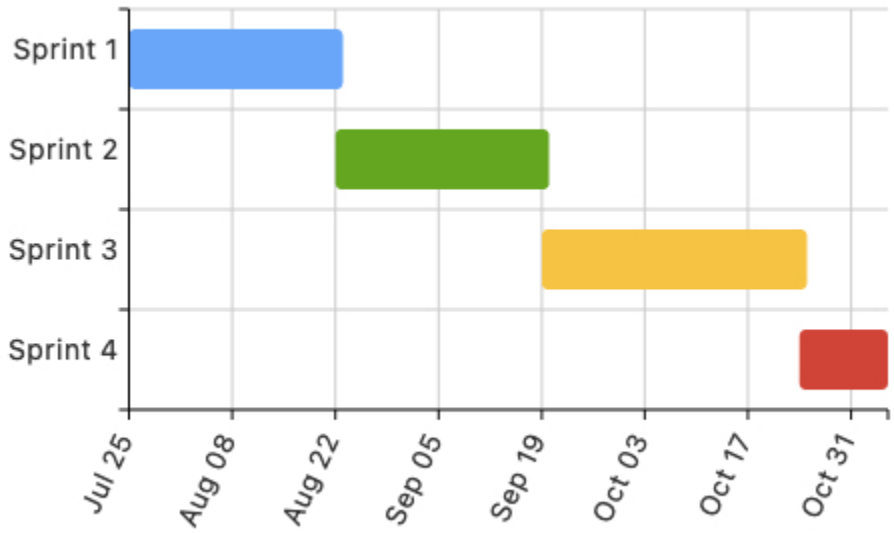
To avoid this, computer simulations (based on existing databases of known material properties and machine learning algorithms) can be used to compare materials in a much more efficient way. One such tool for this approach is the Python library MatMiner, which allows easy access to ready-made datasets and integrates well with other machine learning Python libraries. However, to use this library, the user must have a substantial level of specialised knowledge in machine learning and programming, which most materials engineering do not have. To help them overcome this, the solution proposed by the client describes a simple and intuitive user interface that would act as a bridge between the user and the MatMiner library in the backend. The vision of this product is to make machine learning methodologies more accessible within the materials science and engineering industry as a whole, minimising the time and financial costs involved and leading to a more efficient industry.

The Client

The client of this project is **Dr Christian Brandl**, who is a senior Mechanical Engineering lecturer at the University of Melbourne and has a Ph.D. in Materials Science & Engineering. Aside from his scholarly work, Dr Brandl also acts as a Materials Consultant to clients who come to him seeking recommendations as to which materials may fit their specific needs, or with questions regarding the properties and characteristics of different materials. Dr Brandl hopes to use the product of this project to demonstrate to his clients the capabilities of machine learning and the possibilities it creates in the materials science field.

Schedule

The project will run over four sprints, with the dates of each outlined below.



Sprint	Start	End
Sprint 1	Jul 25	Aug 22
Sprint 2	Aug 22	Sep 19
Sprint 3	Sep 19	Oct 24
Sprint 4	Oct 24	Nov 4

| Personas

Persona #1 - Prepared by team RedBack

Version 1.2 - Final

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."

Alex has a Masters degree in Material Science. He has been working as a Materials Engineer for 15 years, he has been promoted to a senior role over time. Because physical experiments can take years, his job requires him to narrow 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:

Efficiency in finding the right materials
Accuracy of the results
Client satisfaction

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

- Extensible
- Accuracy
- Reliability
- Responsive
- Scalable
- Transparent

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

Version 1.1

Version 1.0

Persona #2 - Prepared by team BoxJelly

Elaine Kim

age: 22
residence: Melbourne
education: Material Engineering
occupation: Postgraduate student
marital status: Single

"It's SO time spending 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

INTERNET

SOFTWARE

MOBILE APPS

SOCIAL NETWORK

Criteria For Success:

Design a usable online tool which Assol can easily perform Materials data requests and retrievals and on a simple click of a button can predict the property of a material with features and ML algorithms chosen by her

Needs

- Easy-to-use Material science data processing and retrieval interface application
- A tool to predict property of a material with ML methods
- A powerful tool or technology supports material researching
- Get a HD on research assignment in each semester

Wants

- A data mining application that helps her research
- 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
- High information density
- Flexible
- Understandable

Fears

- Spends hours without getting anything done because she neither has a adequate tool to do data mining, nor the programming skill to analyze the data herself.
- Looks at the generated diagram without knowing what it means
- Conducts countless experiments just to figure out the properties of the materials
- School assignments are strongly limited by time.
- Hard to choose suitable ML algorithms.

Persona #3 - Prepared by team 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 in my work, but only after it shows its efficiency and reliability to me. 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 every day, 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. Although his company provides some solutions, but they are awkward and only have limited functions.

Comfort With Technology

INTERNET

SOFTWARE

MOBILE APPS

SOCIAL NETWORK

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 and energy 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 about properties he interested in which can be modified with interface
- Help him find the material with best predict properties

Wants

- Ability to interact with the graph to further compare serval materials in detail
- Explain what ML method the system applied and how it helps the prediction

Values

- Easy to get started
- Efficient back-end process
- Abilities to select functions and filter results

Fears

- Frequently unavailability which may waste him much of time to retry
- Not enough guidance in the web or tool so him my feel confused to find functions he wants.

- Well organized visualization of interface and graphs

- No choice of properties that he needs
- Lacking understand of what the system done, which may influence the confidence level of his report

| Motivational Model

- Do/Be/Feel lists
- Goal Model

Do/Be/Feel lists

Version 1.1 - Sprint 1 Final (cross-team effort)

Who (users)	Do (functional goals)	Be (qualitative goals)	Feel (emotional goals)
Students	Add more database, machine learning method and plot types	Accessible	Accomplished
Administrators	Compare data using tables & plots	Accurate	Comfortable
Professionals	Data Pre-processing: Calculate descriptive statistics	Approachable	Confident
Industry Partners	Data Pre-processing: Consider anonymized data	Communicative	Convenient
Teachers	Data Pre-processing: Overview of the current import data	Educational	Curious
Researcher	Data Pre-processing: Reduces noise and eliminates ambiguity	Extensible	Effective
Code maintainers	Data Pre-processing: Standardizing data to bring it into the formatting range	Informative	Empowered
	Data Visualization: Data processing: Tabular data & Plotted Graph	Interactive	Engaged
	Edit python code directly in the interface	Intuitive	Guided
	Export input data	Learnable	Inspired
	Export jupyter notebook file	Legally Compliant	Motivated
	Export output data tables and figures	Reliable	Productive
	Featurization data: Add multiple composition-based features	Responsive	Safe
	Featurization data: Add multiple simple density features	Scalable	Satisfied
	Import Data: Creat working spaces when importing	Secure	
	Import Data: Drag and drop import of files	Transparent (progress, error messages, notebook export...)	
	Import Data: Import data files (CSV, XES, Parquet) from local system		
	Log in/Log out		
	Machine Learning: Define input data and output data: Splitting data into training, test, and validation sets		
	Machine Learning: Determining model features and training the model: Configure and adjust hyperparameters for optimum performance		
	Machine Learning: Evaluate model performance and establish benchmarks: Continuous measurement and monitoring of model performance		
	Machine Learning: Evaluate model performance and establish benchmarks: Evaluate models using validation methods and validation datasets		
	Machine Learning: Get model results: The most important features of the current ML model		
	Machine Learning: Select the machine learning model to be used		
	Maintain software		
	save/load workflows		
	Sign up		

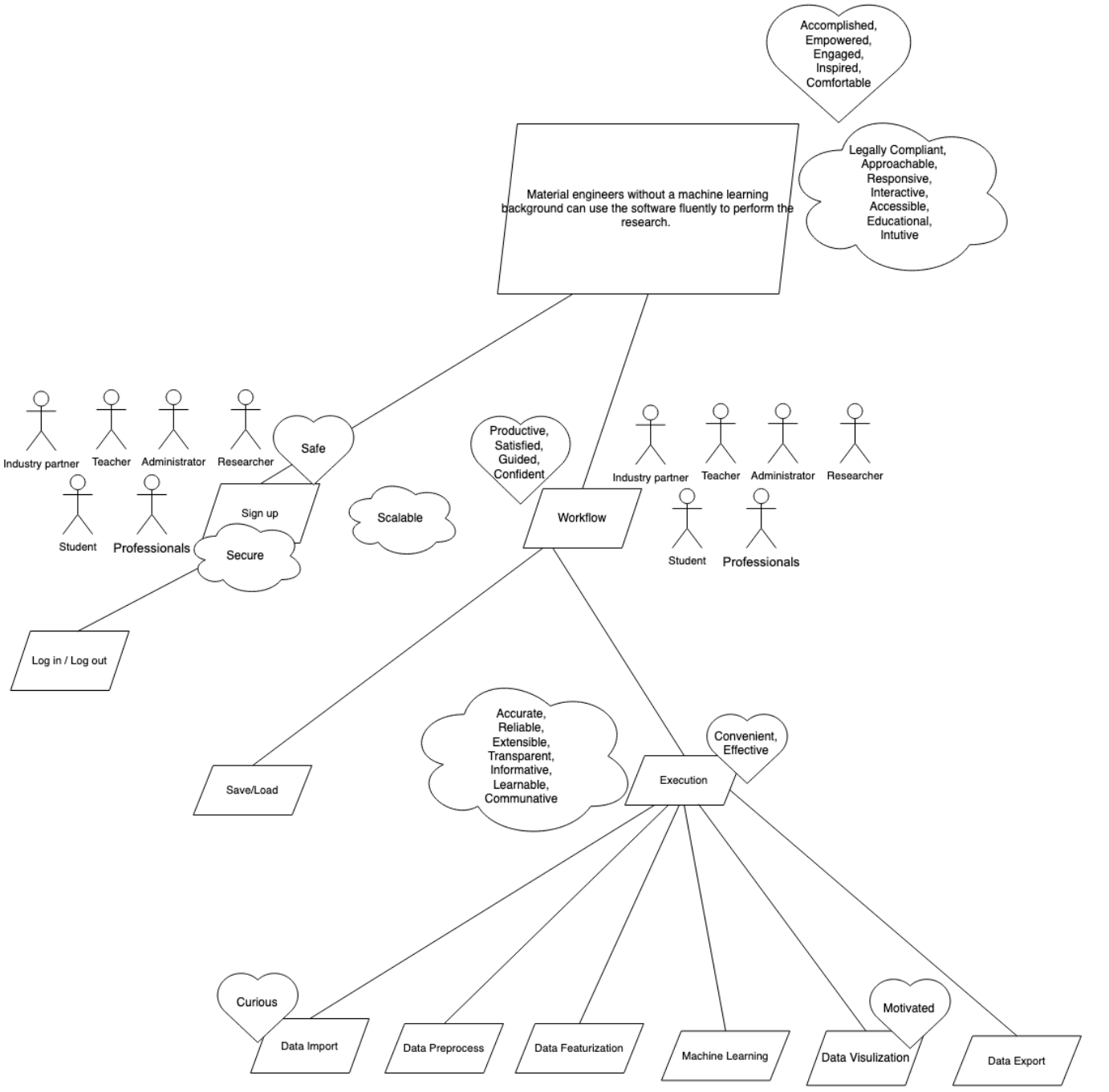
Version 1.0 - Prepared by team RedBack

Who will use it?	With the product, the user can Do	The product should Be	The user should Feel	Concerns (assuming this means "things the product should not do/be/feel")
Technical users	Build and run a basic workflow	Approachable	Empowered	Should not require the user to download anything
Professional users	Build and run an advanced workflow	Reliable	Engaged	Should not require much (any?) python or programming knowledge
Graduate students	Export / fine-tune in notebook	Scalable	Curious	Should require only minimal (wikipedia-level) knowledge of machine learning

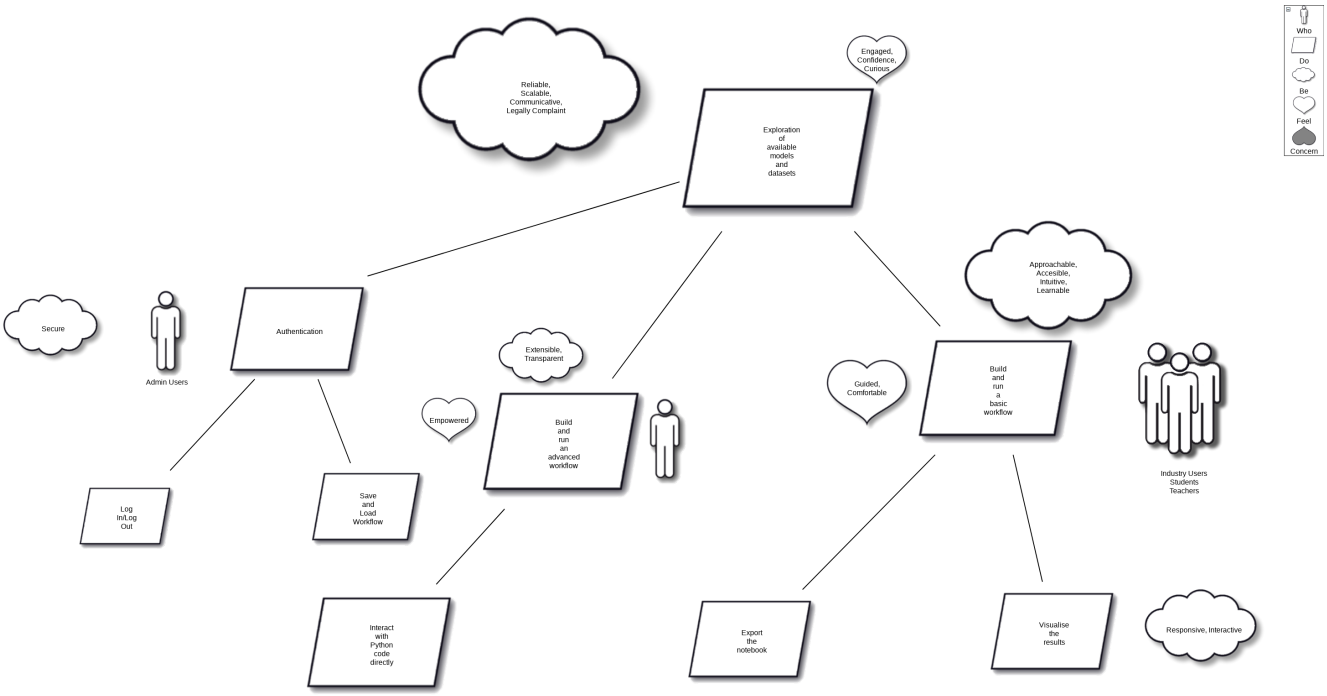
Teachers	Exploration of available models / datasets	Extensible	Confident	
Maintainers	Stop / resume session	Transparent (progress, error messages, notebook export...)	Comfortable	
Admin	Maintain software	Intuitive	Guided	
	Interact with Python code directly	Accessible		
	Visualise data	Secure		
	View relevant citations	Responsive		
	Log In sessions	Learnable		
		Communicative		
		Interactive		
		Legally Compliant		

Goal Model

Version 1.1 - Sprint 1 Final (cross-team effort)



Version 1.0 - Prepared by team RedBack


Who
Do
Be
Feel
Concern
Reliable,
Scalable,
Communicative,
Legally Complaint
Engaged,
Confidence,
Curious
Exploration
of
available
models
and
datasets
Approachable,
Accesible,
Intuitive,
Learnable
Secure
Admin Users
Authentication
Extensible,
Transparent
Empowered
Build
and
run
an
advanced
workflow
Guided,
Comfortable
Build
and
run
a
basic
workflow
Industry Users
Students
Teachers
Log
In/Log
Out
Save
and
Load
Workflow
Interact
with
Python
code
directly
Export
the
notebook
Visualise
the
results
Responsive, Interactive

| User Stories

- Version 1.3 - Sprint 1 Final (cross-team effort)
  - Epics & Owning team allocation
  - User Stories
- Version 1.2 - Revised epics & different size and priority formats (contains additional notes)
- Version 1.1 - Merged between teams
- Version 1.0 - Prepared by team RedBack

Version 1.3 - Sprint 1 Final (cross-team effort)

Epics & Owning team allocation

	Epic	Total Size	Highest Priority within Epic	Assigned Team
1	**Input Data**	27	1 - Must Have	RedBack
2	**Administration**	31	1 - Must Have	BoxJelly
3	**Machine Learning**	16	1 - Must Have	BlueRing
4	**Data Visualisation**	13	1 - Must Have	BlueRing
5	**Jupyter Notebook**	20	2 - Should Have	Unassigned - stretch goal
6	**External Data**	8	2 - Should Have	Unassigned - stretch goal

User Stories

		Role		Action		Goal	Epic	Size (days)	Priority
1	**As a**	general user	**I want to**	be able to view the citations for used featurizers	**so that**	I could be know more about the source of the featurizer (legally compliant)	Input Data	1	1 - Must have
2	**As a**	general user	**I want to**	browse built-in featurizers	**so that**	I can discover ways of manipulating my data		1	1 - Must have
3	**As a**	general user	**I want to**	browse built-in datasets	**so that**	I can discover data to experiment with		1	1 - Must have
4	**As a**	student	**I want to**	quickly browse the Materials available in the database for retrieval and simulations	**so that**	I can quickly perform queries.		3	1 - Must have
5	**As a**	general user	**I want to**	be able to select datasets from exisiting databases	**so that**	I do not have to worry about how the data is loaded		3	1 - Must have
6	**As a**	general user	**I want to**	be able to preview the input data	**so that**	I could explore the data		1	2 - Should have
7	**As a**	general user	**I want to**	Select specific features from a dataset	**so that**	I can improve the precision of my model		3	2 - Should have
8	**As a**	Pro user	**I want to**	add new features	**so that**	they can be reused in the future		5	2 - Should have
9	**As a**	general user	**I want to**	be able to reference / view citation for original data sources	**so that**	I can retrieve data.		1	3 - Could have
10	**As a**	pro user	**I want to**	be able to apply new featurizers	**so that**	I can create new features		3	3 - Could have
11	**As a**	student	**I want to**	clean and tune data input	**so that**	I have less noise on visualizations.		5	3 - Could have
12	**As a**	student	**I want to**	save project specific data/checkpoints	**so that**	I can pick up where I left off for specific projects	Administration	1	1 - Must have

13	**As a**	pro user	**I want to**	export model selections, parameters, and data flows	**so that**	I can save my work and share it with others		1	1 - Must have
14	**As a**	pro user	**I want to**	import exported model selections, parameters, and data flows	**so that**	I can continue work I had previously saved		1	1 - Must have
15	**As a**	student	**I want to**	Create an account using single-sign on, restricted to the *.unimelb.edu.au domain	**so that**	my research remains private		3	1 - Must have
16	**As a**	pro user	**I want to**	Control job execution	**so that**	I can start, view progress of, and cancel jobs related to my project		3	1 - Must have
17	**As a**	pro user	**I want to**	be able to opt in to pro-user features	**so that**	I can access pro user features		5	1 - Must have
18	**As a**	pro user	**I want to**	have my pro user settings persist on each visit	**so that**	I don't have to reconfigure settings to use the features I need		1	2 - Should have
19	**As a**	student	**I want to**	recieve provided hints and guidance for new users	**so that**	I can quickly learn how to use software		3	2 - Should have
20	**As a**	pro user	**I want to**	easily find and read documentation on the pro features	**so that**	I can use them with ease		5	2 - Should have
21	**As a**	pro user	**I want to**	Be kept informed about job status	**so that**	I can avoid polling my workspace to check for results		3	3 - Could have
22	**As a**	pro user	**I want to**	have access to more processing power	**so that**	I can run more complex operations or use more data		5	3 - Could have
23	**As a**	general user	**I want to**	able to select a Machine Learning model	**so that**	I could use it to train and run the data	Machine Learning	1	1 - Must have
24	**As a**	general user	**I want to**	browse built-in ML models	**so that**	I can discover ways of manipulating my data		1	1 - Must have
25	**As a**	user	**I want to**	be able to select split ratio of data	**so that**	to train and test the model		1	2 - Should have
26	**As a**	pro user	**I want to**	have the option to change the hyperparameters used in the machine learning model	**so that**	I can fine tune my test results.		3	2 - Should have
27	**As a**	pro user	**I want to**	be able use additional ML models	**so that**	I can improve accuracy		5	2 - Should have
28	**As a**	pro user	**I want to**	combine multiple models together	**so that**	I can model more complex data manipulations		5	3 - Could have
29	**As a**	general user	**I want to**	see clear annotation or explanation of data points and features	**so that**	I can understand the results of the analysis	Data Visualisation	3	1 - Must have
30	**As a**	student	**I want to**	use different type of plotting graphs	**so that**	I have flexibility to visualize data according to my needs.		5	1 - Must have
31	**As a**	general user	**I want to**	able to view and plot the results of the model	**so that**	I could analysis and visualise the effects of the model		5	1 - Must have
32	**As a**	student	**I want to**	export my work to a Jupyter Notebook	**so that**	I can extend my work beyond the capability of the application	Jupyter Notebook	5	2 - Should have
33	**As a**	general user	**I want to**	attach comments to workflow objects	**so that**	I can document my work		5	3 - Could have
34	**As a**	Pro user	**I want to**	edit python code on the interface	**so that**	I can have control how the ML algorithms works		5	3 - Could have
35	**As a**	Pro user	**I want to**	upload my own script (in python) if possible	**so that**	I can extend the tool to support custom models and featurizers		5	3 - Could have
36	**As a**	pro user	**I want to**	be able to access new databases	**so that**	I can access addional data	External Data	3	2 - Should have
37	**As a**	Pro user	**I want to**	be able to add new datasets in the future	**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

38	**As a**	student	**I want to**	analyze the relationship between different features	**so that**	I can identify which features I need to select for my analysis	?	5	2 - Should have
39	**As a**	general user	**I want to**	add specific materials to the workflow for analysis	**so that**	compare the performance of the specific material my client or I choose with other material	?	3	3 - Could have

Version 1.2 - Revised epics & different size and priority formats (contains additional notes)

		Role		Action	Epic	Domain		Goal	Size (days)	Priority
1	**As a**	general user	**I want to**	be able to view the citations for used featurizers	Input Data	Documentation	**so that**	I could be know more about the source of the featurizer (legally compliant)	1	1 - Must have
2	**As a**	general user	**I want to**	browse built-in featurizers	Input Data	ML	**so that**	I can discover ways of manipulating my data	1	1 - Must have
3	**As a**	general user	**I want to**	browse built-in datasets	Input Data	Server	**so that**	I can discover data to experiment with	1	1 - Must have
4	**As 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
5	**As a**	general user	**I want to**	be able to select datasets from exisiting databases	Input Data	Data Manipulation	**so that**	I do not have to worry about how the data is loaded	3	1 - Must have
6	**As a**	general 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
7	**As a**	general user	**I want to**	Select specific features from a dataset	Input Data	Data Manipulation	**so that**	I can improve the precision of my model	3	2 - Should have
8	**As 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
9	**As a**	general user	**I want to**	be able to reference / view citation for original data sources	Input Data	Documentation	**so that**	I can retrieve data.	1	3 - Could have
10	**As 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
11	**As a**	student	**I want to**	clean and tune data input	Input Data	Data Manipulation	**so that**	I have less noise on visualizations.	5	3 - Could have
12	**As a**	student	**I want to**	save project specific data/checkpoints	Administration	Server	**so that**	I can pick up where I left off for specific projects	1	1 - Must have
13	**As a**	pro user	**I want to**	export model selections, parameters, and data flows	Administration	Server	**so that**	I can save my work and share it with others	1	1 - Must have
14	**As a**	pro user	**I want to**	import exported model selections, parameters, and data flows	Administration	Server	**so that**	I can continue work I had previously saved	1	1 - Must have
15	**As a**	student	**I want to**	Create an account using single-sign on, restricted to the *.unimelb.edu.au domain	Administration	Server	**so that**	my research remains private	3	1 - Must have
16	**As a**	pro user	**I want to**	Control job execution	Administration	Server	**so that**	I can start, view progress of, and cancel jobs related to my project	3	1 - Must have
17	**As a**	pro user	**I want to**	be able to opt in to pro-user features	Administration	UI	**so that**	I can access pro user features	5	1 - Must have
18	**As a**	pro user	**I want to**	have my pro user settings persist on each visit	Administration	Server	**so that**	I don't have to reconfigure settings to use the features I need	1	2 - Should have
19	**As a**	student	**I want to**	recieve provided hints and guidance for new users	Administration	UI	**so that**	I can quickly learn how to use software	3	2 - Should have
20	**As a**	pro user	**I want to**	easily find and read documentation on the pro features	Administration	Documentation	**so that**	I can use them with ease	5	2 - Should have

21	**As a**	pro user	**I want to**	Be kept informed about job status	Administration	Server	**so that**	I can avoid polling my workspace to check for results	3	3 - Could have
22	**As a**	pro user	**I want to**	have access to more processing power	Administration	Server	**so that**	I can run more complex operations or use more data	5	3 - Could have
23	**As a**	general 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
24	**As a**	general 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
25	**As a**	user	**I want to**	be able to select split ratio of data	Machine Learning	Data Manipulation	**so that**	to train and test the model	1	2 - Should have
26	**As 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
27	**As 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
28	**As 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
29	**As a**	general user	**I want to**	see clear annotation or explanation of data points and features	Data Visualisation	UI	**so that**	I can understand the results of the analysis	3	1 - Must have
30	**As a**	student	**I want to**	use different type of plotting graphs	Data Visualisation	UI	**so that**	I have flexibility to visualize data according to my needs.	5	1 - Must have
31	**As a**	general user	**I want to**	able to view and plot the results of the model	Data Visualisation	UI	**so that**	I could analysis and visualise the effects of the model	5	1 - Must have
32	**As 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
33	**As a**	general user	**I want to**	attach comments to workflow objects	Jupyter Notebook	UI	**so that**	I can document my work	5	3 - Could have
34	**As 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
35	**As 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
36	**As a**	pro user	**I want to**	be able to access new databases	External Data	Data Manipulation	**so that**	I can access addional data	3	2 - Should have
37	**As a**	Pro user	**I want to**	be able to add new datasets in the future	External Data	Data Manipulation	**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
38	**As 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
39	**As a**	general user	**I want to**	add specific materials to the workflow for analysis	?	Data Manipulation	**so that**	compare the performance of the specific material my client or I choose with other material	3	3 - Could have

Version 1.1 - Merged between teams

		Role		Action	Epic		Goal	Size	Priority
1	**As a**	student	**I want to**	clean and tune data input	Data Manipulation	**so that**	I have less noise on visualizations.	large	Could have
2	**As a**	general user	**I want to**	attach comments to workflow objects	UI	**so that**	I can document my work	large	Could have
3	**As a**	Pro user	**I want to**	be able to add new datasets in the future	Data Manipulation	**so that**	if there's a new dataset that can be used on a new project, it can be added instantly	large	Could have
4	**As a**	Pro user	**I want to**	edit python code on the interface	UI	**so that**	I can have control how the ML algorithms works	large	Could have

5	**As a**	Pro user	**I want to**	upload my own script (in python) if possible	Server	**so that**	I can extend the tool to support custom models and featurizers	large	Could have
6	**As a**	pro user	**I want to**	have access to more processing power	Server	**so that**	I can run more complex operations or use more data	large	Could have
7	**As a**	pro user	**I want to**	combine multiple models together	ML	**so that**	I can model more complex data manipulations	large	Could have
8	**As a**	student	**I want to**	use different type of plotting graphs	UI	**so that**	I have flexibility to visualize data according to my needs.	large	Must have
9	**As a**	general user	**I want to**	able to view and plot the results of the model	UI	**so that**	I could analysis and visualise the effects of the model	large	Must have
10	**As a**	pro user	**I want to**	be able to opt in to pro-user features	UI	**so that**	I can access pro user features	large	Must have
11	**As 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	large	Should have
12	**As a**	student	**I want to**	export my work to a Jupyter Notebook	Server	**so that**	I can extend my work beyond the capability of the application	large	Should have
13	**As a**	Pro user	**I want to**	add new features	ML	**so that**	they can be reused in the future	large	Should have
14	**As a**	pro user	**I want to**	easily find and read documentation on the pro features	Documentation	**so that**	I can use them with ease	large	Should have
15	**As a**	pro user	**I want to**	be able use additional ML models	ML	**so that**	I can improve accuracy	large	Should have
16	**As a**	general user	**I want to**	add specific materials to the workflow for analysis	Data Manipulation	**so that**	compare the performance of the specific material my client or I choose with other material	medium	Could have
17	**As a**	pro user	**I want to**	Be kept informed about job status	Server	**so that**	I can avoid polling my workspace to check for results	medium	Could have
18	**As a**	pro user	**I want to**	be able to apply new featurizers	ML	**so that**	I can create new features	medium	Could have
19	**As a**	student	**I want to**	quickly browse the Materials available in the database for retrieval and simulations	UI	**so that**	I can quickly perform queries.	medium	Must have
20	**As a**	student	**I want to**	Create an account using single-sign on, restricted to the *.unimelb.edu.au domain	Server	**so that**	my research remains private	medium	Must have
21	**As a**	general user	**I want to**	be able to select datasets from exisiting databases	Data Manipulation	**so that**	I do not have to worry about how the data is loaded	medium	Must have
22	**As a**	general user	**I want to**	see clear annotation or explanation of data points and features	UI	**so that**	I can understand the results of the analysis	medium	Must have
23	**As a**	pro user	**I want to**	Control job execution	Server	**so that**	I can start, view progress of, and cancel jobs related to my project	medium	Must have
24	**As a**	student	**I want to**	recieve provided hints and guidance for new users	UI	**so that**	I can quickly learn how to use software	medium	Should have
25	**As a**	general user	**I want to**	Select specific features from a dataset	Data Manipulation	**so that**	I can improve the precision of my model	medium	Should have
26	**As a**	pro user	**I want to**	have the option to change the hyperparameters used in the machine learning model	ML	**so that**	I can fine tune my test results.	medium	Should have
27	**As a**	pro user	**I want to**	be able to access new databases	Data Manipulation	**so that**	I can access addional data	medium	Should have
28	**As a**	general user	**I want to**	be able to reference / view citation for original data sources	Documentation	**so that**	I can retrieve data.	small	Could have
29	**As a**	student	**I want to**	save project specific data/checkpoints	Server	**so that**	I can pick up where I left off for specific projects	small	Must have

30	**As a**	general user	**I want to**	be able to view the citations for used featurizers	Documentation	**so that**	I could be know more about the source of the featurizer (legally compliant)	small	Must have
31	**As a**	general user	**I want to**	able to select a Machine Learning model	ML	**so that**	I could use it to train and run the data	small	Must have
32	**As a**	general user	**I want to**	browse built-in featurizers	ML	**so that**	I can discover ways of manipulating my data	small	Must have
33	**As a**	general user	**I want to**	browse built-in ML models	ML	**so that**	I can discover ways of manipulating my data	small	Must have
34	**As a**	general user	**I want to**	browse built-in datasets	Server	**so that**	I can discover data to experiment with	small	Must have
35	**As a**	pro user	**I want to**	export model selections, parameters, and data flows	Server	**so that**	I can save my work and share it with others	small	Must have
36	**As a**	pro user	**I want to**	import exported model selections, parameters, and data flows	Server	**so that**	I can continue work I had previously saved	small	Must have
37	**As a**	general user	**I want to**	be able to preview the input data	UI	**so that**	I could explore the data	small	Should have
38	**As a**	pro user	**I want to**	have my pro user settings persist on each visit	Server	**so that**	I don't have to reconfigure settings to use the features I need	small	Should have
39	**As a**	user	**I want to**	be able to select split ratio of data	Data Manipulation	**so that**	to train and test the model	small	Should have

Version 1.0 - Prepared by team RedBack

Strikethrough rows were identified as duplicates after the user stories created by individual team members were all merged.

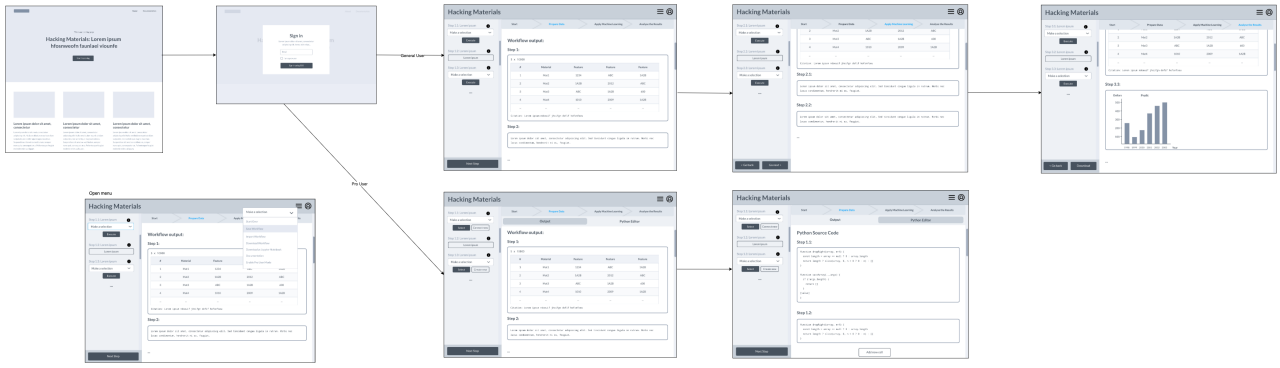
		Role		Action		Goal
1	**As a**	admin	**I want to**	assign different user types	**so that**	I can allow pro users to access more functionalities
2	**As a**	admin	**I want to**	be able to cancel a running process	**so that**	I can free up resources for more urgent tasks
3	**As a**	general user	**I want to**	have hints for the next action popped up	**so that**	I do not have to worry about technical details of the workflow
4	**As a**	general user	**I want to**	know which are the most important features	**so that**	I can use them to improve the performance of the model
5	**As a**	general user	**I want to**	be able to select datasets from exisiting databases	**so that**	I do not have to worry about how the data is loaded
6	**As a**	general user	**I want to**	be able to preview the data	**so that**	I could explore the data or clean it
7	**As a**	general user	**I want to**	perform feature selection / column deletion	**so that**	I could have access to the subset of data needed
8	~~**As a**~~	~~general user~~	~~**I want to**~~	~~be able to access the web interface~~	~~**so that**~~	~~I could perform all the exploration tasks~~
9	**As a**	general user	**I want to**	be able to view the citations for used featurizers	**so that**	I could be know more about the source of the featurizer (legally compliant)
10	**As a**	general user	**I want to**	able to select the columns needed for input and output	**so that**	I could use it to train the Machine Learning model
11	**As a**	general user	**I want to**	able to select a Machine Learning model	**so that**	I could use it to train and run the data

12	**As a**	general user	**I want to**	able to view and plot the results of the model	**so that**	I could analysis and visualise the effects of the model
13	**As a**	general user	**I want to**	browse built-in featurizers and ML models	**so that**	I can discover ways of manipulating my data
14	**As a**	general user	**I want to**	browse built-in datasets	**so that**	I can discover data to experiment with
15	**As a**	general user	**I want to**	attach comments to workflow objects	**so that**	I can document my work
16	~~**As a**~~	~~general user~~	~~**I want to**~~	~~have user-friendly access to the interface~~	~~**so that**~~	~~I do not lose my interest in doing any operation~~
17	**As a**	general user	**I want to**	have secure access to the interface	**so that**	I have the assurity that my data is in a safe place
18	~~**As a**~~	~~pro user~~	~~**I want to**~~	~~interact with python code directly~~	~~**so that**~~	~~I can customise my workflow further~~
19	**As a**	pro user	**I want to**	apply featurizers	**so that**	I have more options for my ML model
20	**As a**	pro user	**I want to**	connect to external databases	**so that**	I can use the data that satisfies my complex user case
21	**As a**	pro user	**I want to**	have the option to select more model types	**so that**	I can be more confident in the results when accuracy is important
22	**As a**	pro user	**I want to**	have the option to select more plot types	**so that**	I can visualise the results in ways that suit my more complex needs
23	**As a**	pro user	**I want to**	have my pro user settings persist on each visit	**so that**	I don't have to reconfigure settings to use the features I need
24	**As a**	pro user	**I want to**	easily find and read documentation on the pro features	**so that**	I can use them with ease
25	**As a**	pro user	**I want to**	have access to more processing power	**so that**	I can run more complex operations or use more data
26	~~**As a**~~	~~pro user~~	~~**I want to**~~	~~save the output of my workflow~~	~~**so that**~~	~~I can leave long operations running without worrying about disconnecting~~
27	**As a**	pro user	**I want to**	have the option to obtain statistics of the data	**so that**	I get insightful information about the features
28	~~**As a**~~	~~pro user~~	~~**I want to**~~	~~edit model (hyper-)parameters~~	~~**so that**~~	~~I can experiment with non-default settings for model behaviour~~
29	**As a**	pro user	**I want to**	combine multiple models together	**so that**	I can model more complex data manipulations
30	**As a**	pro user	**I want to**	export model selections, parameters, and data flows	**so that**	I can save my work and share it with others
31	**As a**	pro user	**I want to**	import exported model selections, parameters, and data flows	**so that**	I can continue work I had previously saved
32	**As a**	pro user	**I want to**	generate a Jupyter notebook from my workspace	**so that**	I can access the greater degree of control and flexibility offered by a programming interface
33	**As a**	pro user	**I want to**	browse known online datasets	**so that**	I can access data required for my work

34	**As a**	pro user	**I want to**	derive new features from existing features	**so that**	I can provide richer descriptions of my data
35	**As a**	pro user	**I want to**	access custom databases, featurizers, and ML models	**so that**	I can use data and code I have developed elsewhere
36	**As a**	pro user	**I want to**	Control job execution	**so that**	I can start, view progress of, and cancel jobs related to my project
37	**As a**	pro user	**I want to**	Persist workspace objects and status	**so that**	I can close my workspace without losing my work
38	**As a**	pro user	**I want to**	Be kept informed about job status	**so that**	I can avoid polling my workspace to check for results
39	**As a**	pro user	**I want to**	have the option to modify the python code in the user interface	**so that**	I can modify the code directly to achieve the functionality I want.
40	**As a**	pro user	**I want to**	add new features to the existing features list	**so that**	I can include more useful features for materials experiments.
41	**As a**	pro user	**I want to**	remove selected features from the existing feature list	**so that**	I can remove features that are irrelevant to the experiments.
42	~~**As a**~~	~~pro user~~	~~**I want to**~~	~~be able to specify any online database I want to use~~	~~**so that**~~	~~I can use data from different sources.~~
43	**As a**	pro user	**I want to**	have the option to change the hyperparameters used in the machine learning model	**so that**	I can fine tune my test results.
44	~~**As a**~~	~~pro user~~	~~**I want to**~~	~~be informed when my running task is done~~	~~**so that**~~	~~I can view my results in time~~

| Low-fidelity Prototype

The prototype was created in Marvel and can be viewed here.



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 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. 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 the 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.

| Notes on provided material

Organised Wishlist

As described in the provided notebook and the rough prioritisation described in the kickoff meeting.

High priority features:

- Save, load and modify workflow within interface
- Web interface with user logins
- Server based online application
- Documentation **- Assumed**

Low priority features:

- Create new features based on existing features (PRO user?)
- Interface to interact with python code directly (PRO interface)
- **Extendable** to include more databases, machine learning methods and plot types (e.g., by PRO user)
- Hints on sensible/possible next steps and interaction with UI
- List all required citation </aside>

Unknown priority features:

- Define a set of user defined features
- Generates corresponding python code (or jupyther file with markdown)
- Download and upload files to/from server (jupyther file?)
- Interaction with visualisation, e.g.,click on data point in plot and open corresponding data entry </aside>

General characteristic:

- Should be usable by engineer or postgrad engineering students without programming experience and just WIKIPEDIA knowledge of machine learning

MatMiner example workflow Requirements

Key
Not described by the client, just a suggestion from within the team
Requirement suggested in the relevant step within the Jupyter notebook
Requirement included in the wishlist

#	Workflow step	Requirement
1.0 A	Select dataset from built in datasets with *matminer.dataset*	UI to select dataset from available options
1.0 B	Select a dataset from an online resource with *matminer. data_retrieval*	UI to set up the connection to the online databases
1.1	Preview DataFrame to explore or clean it	“Preview data” button prints same result?
1.2	Remove unneeded columns	Allow user to remove columns from list?
1.3	Review descriptive statistics of the data with *describe()*	“Describe” button prints same result?
2.1, 2.2, 2.3	Featurization: convert inputs into numbers that meaningfully represent the underlying physical quantity using the descriptors library	Select from possible features available
2.1	Printing citation for each featurizer using *citations()*	List required citation (. *citation()* result) for all used featurizers
3.1	Define input and output data of the ML model	Select output column from list of columns? All remaining numerical featurized data assumed as input?
3.2.1	Pick, train and run the ML model to generate predictions	Select from list of possible models?
3.2.2	Display the fitting score and cross-validation score	Automatically print results after predictions?
3.2.3	Plot results using plotting library	Select plot type? Select plotting library? Interaction with visualisation (e.g.,click on data point in plot and open corresponding data entry)
3.3.1	Repeat 3.2.1 with different models	Save previous results and allow to run more models?
3.3.2	Plot training error instead of test error (optionally)	Pick which data to plot?
3.3.3	See which features were most important in the model	“Analyse” option?

| Development

- | Development Process

| Development Process

- 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 more granularity
        - Code reviews:
            - Require tests 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

| Assessments

This section covers additional assessable deliverables that are not part of the project documentation to be submitted to the client.

- | Milestones & Deliverables
- | Cross-Team Collaboration Business Case
- | Reflections

| Cross-Team Collaboration Business Case

Formatted document can be downloaded here: HA Software Project - Cross-team collaboration - Business case.pdf

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:

- 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 follow:

- 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 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 their 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

- **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 3pm.
- **Shared GitHub repository:** If our proposal is accepted, we would need a shared GitHub repository so that Dr. Brandl will not have three copies of each deliverable.

| Resources

- | Links and Tools
- | Contact Information

| Links and Tools

- GitHub organisation: https://github.com/orgs/COMP90082-2022-SM2/teams
- GitHub team: https://github.com/orgs/COMP90082-2022-SM2/teams/ha-redback
- Confluence: https://confluence.cis.unimelb.edu.au:8443/display/COMP900822022SM2HARedBack/Home
- When2Meet: https://www.when2meet.com
    - Our pre-filled one based on our regular schedules: https://www.when2meet.com/?16293461-QLKH3
- Trello board: https://trello.com/b/jZLcqPMi/agile-sprint-board
- Other tools not yet set up.
- https://edstem.org/au/courses/8857/discussion/
- This Person Does Not Exist: https://thispersondoesnotexist.com/
- Persona design tools: Xtensio: https://xtensio.com/, UXpressia: https://uxpressia.com/

- https://www.conventionalcommits.org/en/v1.0.0/ for a guide on writing commit messages
- https://www.atlassian.com/git/tutorials/comparing-workflows/gitflow-workflow for a guide on "Git Flow", a branching strategy

| Contact Information

Name	Calling name	Email	GitHub	Student ID	Discord
Alastair Daivis (Team rep)	Alastair	adaivis@student.unimelb.edu.au	alaroldai	359 101	alaroldai
Sanjeevani Avasthi	Sanjee	savasthi@student.unimelb.edu.au	sanjeevania	1101265	sanjeevani
Mamta Lopes	Mamta	mlopes@student.unimelb.edu.au	mamtal	1157314	Mamta
Ghina Yashar	Ghina	gyashar@student.unimelb.edu.au	GhinaY	1274878	Ghinaenae
Chunbaixue Yang	Caitlyn	chunbaixue@student.unimelb.edu.au	ycbx1993	1208333	KitKat

Client

Name	Calling name	Email	GitHub
Dr Christian Brandl		christian.brandl@unimelb.edu.au	

Staff

Name	Calling name	Email	GitHub
Mauro Mello Jr (supervisor)		mauro.mellojr@unimelb.edu.au	
Eduardo Oliviera		eduardo.oliveira@unimelb.edu.au	agogear

Sibling team contacts: BlueRing

Name	Calling name	Email	GitHub	Student ID	Discord
Yanan Liu (Team rep)		yananl7@student.unimelb.edu.au			
Hongpeil Lu		hongpeil@student.unimelb.edu.au			
Jiahao Ju		jiahaoj1@student.unimelb.edu.au			
Xinle Yu		xinley@student.unimelb.edu.au			
Rui Zhang		rzzhan2@student.unimelb.edu.au			

Sibling team contacts: BoxJelly

Name	Calling name	Email	GitHub	Student ID	Discord
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			