ML Experiment Management Tools: A Usability & Learnability Study

In this experiment, you will perform a series of tasks with ML experiment management tools. These tools are supposed to assist you in tracking and organizing the assets required and generated when running multiple iterations of ML experiments. The main goal of this experiment is to evaluate usability, learnability, and the level of support offered by these tools. In this experiment, you will be asked to carry out some guided ML tasks while using two tools—Neptune.ai and DVC. This experiment has been designed to guide you through some steps of simple regression tasks to reflect a typical data scientist's workflow, which usually involves multiple runs of experiments before arriving at optimal intended results. For example, you will make changes to the datasets, the learning algorithm, and the learning parameters. These steps will be done using support from the two target tools and an additional manual approach with "no-tool."

After being guided through the ML tasks using the subject tools (i.e., Neptune, DVC, and "No-Tool"), you will be asked traceability-related questions to help evaluate how well you are able to track, query, and retrieve the tracked assets. Finally, you will be asked to answer general questions about your experience with the tools.

Participant's Details	
What is your current occupation?	
Student	
What is your level of education?	
B.Sc.	
M.Sc.	
PhD.	
Other:	

How many years of ML experience do you have?
2 months
Do you have any experience with the Git version control tool?
Yes
O No
If your answer is yes, how many years of experience do you have with Git?
4
Do you have any prior experience with any ML experiment tracking tools?
Yes
No
If your applyor is you have the MI experiment tracking to de
If your answer is yes, name the ML experiment tracking tools.
Write down your email address:

Your Task

We have prepared your tasks based on your assigned group.

Select your group *		
Group A		
Group B		
Group C		

Group A

The following tasks are designed for participants assigned to group A only. You will be guided through three different phases in the following order:

- * Phase 1 Neptune
- * Phase 2 DVC
- * Phase 3 No-Tool

Phase 1: Neptune

Neptune.ai is a machine learning experiment tool used to track Machine learning assets such as datasets, parameters, metrics, etc. The tool can be used to track, retrieve and query the assets different runs of an experiment by instrumenting code. The assets are mainly tracked as metadata which can be viewed on a web dashboard for post-experiment analysis. The web dashboard allows users to view the experimental runs that have been done, their results, and associated assets.

Tutorial: Getting started with Neptune

You have already received a link to a brief Neptune tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with the key features of Neptune that are relevant to this experiment.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:

Experiment's Task Follow the instructions provided in the experiment task document for the Group-A Neptune phase. Link to the instructions:
Link to python script:
Questions:
Retrieving & Querying Tracked Data: You will be asked to answer a number of retrieving and querying related questions. You are expected to use the Neptune tool alone to assist you in answering the questions (Do not consult the task instruction document). Try your best to provide the correct answer.
Which of the 13 runs performed best? (i.e. which has the lowest RMSE score?)
What is the RMSE value for that run?
Which of the algorithms (LR & RFR) performed best (lowest RMSE) in their first run?
○ LR
RFR
O I don't know
What data features were used for the experimental run with the highest r2_score?

Compare Run 4 and	Run 1, which one	had the highest m	nean absolute erro	r?
Run 4 Run 1 I don't know				
What was the value	?			
Compare "EX-5" ,"E	X-7", "EX-9" and "	EX-11" in the dashl	poard table.	
	EX-5	EX-7	EX-9	EX-11
Highest RMSE	\circ	\circ	\circ	\circ
Highest R2	\circ	\circ	\circ	\circ
Highest mean absolute error	0			
If we want to reproduce model was used for RandomForestRe	"EX-4"?		e need to retrieve	
	(normalize=False)	1		,
		=10, min_samples st	olit=12, n_estimators	=45)
Other:	Ç (1 <u></u>	,	,	,
Other.				

From the dashboa	ard table fil	ter out the				
Query Neptune for normalize paramet		l with the v	vorst RMSE	(Largest v	alue). Provi	de Run id, and the
List all linear regres	ssion runs v	with RMSE	value less t	han 6.5. Pr	ovide Run i	d, and the r2 value
In the python file, p You are expected to 99			-			
You are expected to	retrieve this		-			
You are expected to	eptune:	information	via the pyth	on script. Yo	ou can write	the code below line
You are expected to 199 Experience with Ne	eptune:	information	via the pyth	on script. Yo	ou can write	the code below line
You are expected to 199 Experience with Ne	eptune:	ptune tutor	rial provide	on script. Yo	this experi	the code below line
You are expected to 199 Experience with New How helpful was the	eptune: 1 O	ptune tutor	rial provide	d ahead of	this experi	ment?
You are expected to 199 Experience with New How helpful was the Not Helpful	eptune: 1 O	ptune tutor	rial provide	d ahead of	this experi	ment?

How helpful is the visual dashboard provided by Neptune.ai when comparing the experimental runs?							
	1	2	3	4	5		
Not Helpful	0	0	0	0	0	Very Helpful	
How long did it tak	e you to co	mplete the	e Neptune.a	ai section?			
0 - 30 mins							
30 mins - 1 hour							
1 hour - 1 hour 3	0 mins						
1 hour 30 mins -	2 hours						
2+ hours							
Any Additional Cor	nments:						
Phase 1 Completed Please proceed to the next phase							
Phase 2: DVC							

Data Version Control is an experiment management tool that extends Git to make ML models shareable and reproducible. It is designed to handle large files, datasets, machine learning models, metrics as well as code.

Tutorial: Getting started with DVC

You have already received a link to a brief DVC tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with the key features of DVC that are relevant to this experiment.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:



Experiment's Task

Follow the instructions provided in the experiment task document for the Group-A DVC phase.

Link to the instructions:

Questions (DVC):

Retrieving & Querying data:

You will asked to perform a number of retrieving tasks, and you will use the tool to assist you in performing these tasks. Try your best to provide the correct answer:

Which run performed best, (i.e. has the lowest RMSE score)?

What is the RMSE value for that run?

Which of the algorithms performed best in their first run? LR or RFR i.e. which one had the lowest RMSE.
○ LR
RFR
O I don't know
What data features were used for the run with the lowest r2_score?
Between Run 4 and Run 1, which one had the highest mean absolute error?
Run 4
Run 1
O I don't know
What was the value?

Compare "Run 5", "Run 7", "Run 9" and "Run 13":								
	Run 5	Run 7	Run 9	Run 13				
Highest RMSE	\bigcirc	\circ	\circ	\circ				
Highest R2	0	\circ	\circ	\circ				
Highest mean absolute error	0	0	0	0				
If we want to reprod model was used for		f previous runs, we	e need to retrieve t	the model. Which				
RandomForestRe	gressor(max_depth:	=5, min_samples_spl	it=6, n_estimators=50	0)				
LinearRegression	(normalize=False)							
LinearRegression	(normalize=True)							
Other:								
Using DVC commands, solve the following:								
Return the run with the worst RMSE (Largest value) Provide tag name, and the parameters used.								
Find the runs that produced models evaluation metric, r2 is greater than 0.32 Provide tag number.								

Vhere you used li				e RIVISE Wa	s greater tr	
Vhat was the R2 v		-				
xperience with D						
low helpful was t	he short tut 1	orial in the 2	beginning 3	in complet	ing the task 5	cs?
Not Helpful	0	0	0	0	0	Very Helpful
low do you rate t	he ease of (completing	the tasks?	,		
	1	2	3	4	5	
Not Easy	0	0	0	0	0	Very Easy
	the DVC co	mmands to	o compare	the runs?		
low helpful were						
low helpful were	1	2	3	4	5	

How long did it take you to complete the DVC section?
O - 30 mins
30 mins - 1 hour
1 hour - 1 hour 30 mins
1 hour 30 mins - 2 hours
2+ hours
Any Additional Comments:
Phase 3: No Tool In this part of the experiment, you will run the experiment with python only. And will you will not have any experiment tools to assist you.
Tutorial: Getting started with No-Tool You have already received a link to a brief No-Tool tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with running the project with No-Tool. We have also provided the link below in case you would like to review it. Follow the link for the tutorial:
Experiment's Task Follow the instructions provided in the experiment task document for the Group-A Neptune phase. Link to the instructions: Link to python script:

Questions (No-Tool):

	Retrieving	&	Querv	ving	data:
--	------------	---	-------	------	-------

You will asked to perform a number of retrieving tasks, and you will use the tool to assist you in performing these tasks. Try your best to provide the correct answer:

Which run performed best, (i.e. has the lowest RMSE score)?
What is the RMSE value for that run?
Which of the algorithms performed best in their first run? LR or RFR i.e. which one had the lowest RMSE.
□ LR□ RFR□ I don't know
What data features were used for the run with the lowest r2_score?
Between Run 4 and Run 1, which one had the highest mean absolute error?
Run 4 Run 1 I don't know

		d "Run 13":		
	Run 5	Run 7	Run 9	Run 13
Highest RMSE	\circ	\circ	\circ	\circ
Highest R2	\circ	0	\circ	\circ
Highest mean absolute error	0	0	0	0
nodel was used for	"Run 4"?	f previous runs, we =5, min_samples_spl		
LinearRegression RandomForestRe	"Run 4"? gressor(max_depth: (normalize=False)		it=6, n_estimators=5	0)
nodel was used for RandomForestRe LinearRegression	"Run 4"? gressor(max_depth: (normalize=False)	=5, min_samples_spl	it=6, n_estimators=5	0)

Vhat was the R2 v	alue of the	very first r	un?			
xperience with 'N	lo-Tool':					
low helpful was t	he short tu	torial in the	beginning	in complet	ing the task	ks?
	1	2	3	4	5	
Not Helpful	0	0	0	0	0	Very Helpful
low do you rate t	he ease of	completing	the tasks?	,		
	1	2	3	4	5	
Not Easy	\bigcirc	0	0	\bigcirc	\bigcirc	Very Easy

How helpful was th	•	our manual	technique	in tracking	/retrieving	the experimental
	1	2	3	4	5	
Not Helpful	0	0	0	0	0	Very Helpful
How long did it tak	e you to co	mplete the	e No-Tool s	ection?		
0-30 mins						
30 mins-1 hour						
1 hour-1 hour 30	mins					
1 hour 30 mins-2	2 hours					
2+ hours						
Any Additional Cor	nments:					
Experience with all	Tools (Nep	tune, DVC	, and 'No-T	ool'		
Even though the exper typical machine learni Consequently, it becor without any supporting	ng task usua nes harder f	ally has man	y more runs	carried out	over a longe	r period of time.
To answer the followir is probably working or						e learning expert who
Describe your expe	erience wit	h each of t	he tools(N	eptune, DV	C and No-	Гооl).

The DVC and Neptune tools provide a significant support for tracking, querying and retrieving generated data from ML experiments over No-tool.
Strongly Agree
Agree
O Neutral
Disagree
Strongly Disagree
Please elaborate on your response above.
Which tool do you consider best for tracking data during ML experiments?
Neptune.ai
O DVC
O No-tool
Which tool do you consider best for querying and retrieving previously tracked data?
Neptune.ai
O DVC
O No-tool

Which of Neptune and DVC do you consider least intrusive in completing the tasks?
Neptune.ai
O DVC
Which of Neptune and DVC was the easiest to learn?
Neptune.ai
O DVC
Which of the tools provides the best support for comparing different experiment runs?
Neptune.ai
O DVC
Neptune helps compare different runs using a web dashboard, while DVC uses CLI. Which do you most convenient?
Neptune (Web dashboards)
O DVC (CLI)
Please elaborate.

Finally, which tool would you recommend a ML practitioner to use?
Neptune.ai
O DVC
O No-Tool
Please elaborate.

Group B:

The following tasks are designed for participants assigned to group B only. You will be guided through three different phases in the following order:

- * Phase 1 DVC
- * Phase 2 No-Tool
- * Phase 3 Neptune.ai

Phase 1: DVC

Data Version Control is an experiment management tool that extends Git to make ML models shareable and reproducible. It is designed to handle large files, datasets, machine learning models, metrics as well as code.

Tutorial: Getting started with DVC

You have already received a link to a brief DVC tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with the key features of DVC that are relevant to this experiment.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:

_				-	
Evi	nar	'im	ant'c	s Tas	ͷ
ᅜᄼ	DEI	1111		o ias	n

Follow the instructions provided in the experiment task document for the Group-A DVC phase.

Link to the instructions:

Questions (DVC):

Retrieving & Querying data:

You will asked to perform a number of retrieving tasks, and you will use the tool to assist you in performing these tasks. Try your best to provide the correct answer:

Which run performed best, (i.e. has the lowest RMSE score)?

Runs: 6 and 7

What is the RMSE value for that run?

58.84

Which of the algorithms performed best in their first run? LR or RFR i.e. which one had the lowest RMSE.

O LR

RFR

O I don't know

What data features were used for the run with the highest r2_score?

's1','sex'

Between Run 4 and	Run 1, which one l	had the highest m	ean absolute error	?
Run 4				
Run 1				
O I don't know				
What was the value	?			
Compare "Run 5", "	Run 7", "Run 9" and	d "Run 13":		
	Run 5	Run 7	Run 9	Run 13
Highest RMSE	0	•	0	0
Highest R2	•	\circ	\circ	\circ
Highest mean absolute error				
If we want to reprod model was used for		f previous runs, we	e need to retrieve	the model. Which
RandomForestRe	gressor(max_depth	=5, min_samples_spl	it=6, n_estimators=5	0)
LinearRegression	n(normalize=False)			
RandomForestRe	egressor(max_depth	=10, min_samples_sp	olit=12, n_estimators	=45)
Other:				

Using DVC comm	ands, solve	the follow	ring:			
Return the run wit Provide tag name, a run 3			gest value)			
Where you used li Provide tag name, a run 6, True, run 7, Tru	nd the param	_	thm and th	e RMSE wa	s less than o	60
What was the R2 v Hint: Use dvc metric 0.07234		-				
Experience with D	VC:					
How helpful was t	he short tut	torial in the	beginning	in complet	ing the task	s?
	1	2	3	4	5	
Not Helpful	0	0	0	0	•	Very Helpful
How do you rate t	he ease of (completing	the tasks?)		
	1	2	3	4	5	
Not Easy	0	0	•	0	0	Very Easy

How helpful were the DVC commands to compare the runs?						
	1	2	3	4	5	
Not Helpful	0	0	0	0	•	Very Helpful
How long did it tak	e you to co	mplete the	e DVC secti	ion?		
0 - 30 mins						
30 mins - 1 hour						
1 hour - 1 hour 3	0 mins					
1 hour 30 mins -	2 hours					
2+ hours						
Any Additional Cor	nments:					
I prefer using CLI to to	rack the asse	ets, easy and	d short comr	mands		

Phase 2: No Tool

In this part of the experiment, you will run the experiment with python only. And will you will not have any experiment tools to assist you.

Tutorial: Getting started with No-Tool

You have already received a link to a brief No-Tool tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with running the project with No-Tool.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:

Experiment's Task Follow the instructions provided in the experiment task document for the Group-A Neptune phase.
Link to the instructions:
Link to python script:
Questions (No-Tool):
Retrieving & Querying data: You will asked to perform a number of retrieving tasks, and you will use the tool to assist you in performing these tasks. Try your best to provide the correct answer:
Which run performed best, (i.e. has the lowest RMSE score)? run 5
What is the RMSE value for that run?
Not sure
Which of the algorithms performed best in their first run? LR or RFR i.e. which one had the lowest RMSE.
○ LR
RFR
I don't know
What data features were used for the run with the lowest r2_score? not sure

Between Run 4 and Run 1, which one had the highest mean absolute error?					
Run 4					
Run 1					
l don't know					
And what was the va	alue?				
not sure					
Compare "Run 5", "F	Run 7", "Run 9" and	d "Run 13":			
	Run 5	Run 7	Run 9	Run 13	
Highest RMSE	•	\circ	\circ	\circ	
Highest R2	\circ	\circ		\circ	
Highest mean absolute error		0	0	0	
	"Run 4"?		e need to retrieve t		
RandomForestRe	gressor(max_depth=	=10, min_samples_sp	olit=12, n_estimators	=45)	
Other:					

For the following write down the run number along with the value:						
Run with the worst Provide the run nr an		rgest value))			
Where you used lir Provide Run number, IDK		_		e RMSE was	s less than o	5.5
What was the R2 v	alue of the	very first r	un?			
Experience with No		torial in the	boginning	in complet	ing the teel	·n 2
How helpful was th						S?
	1	2	3	4	5	
Not Helpful	0	•	0	0	0	Very Helpful
How do you rate th	ne ease of	completing	the tasks?	,		
	1	2	3	4	5	
Not Easy	•	0	0	0	0	Very Easy

Describe how you manually track, query, and retrieve the experiment runs and their assets.						
I did not track the data in any way, which made it difficult to complete the tasks						
How helpful was th		our manual	technique	in tracking	/retrieving	the experimental
	1	2	3	4	5	
Not Helpful	•	0	0	0	0	Very Helpful
How long did it tak	e you to co	mplete the	No-Tool s	ection?		
O-30 mins						
30 mins - 1 hour						
1 hour - 1 hour 30 mins						
1 hour 30 mins	· 2 hours					
2+ hours						
Any Additional Cor	nments:					
Phase 3: Neptune	na laorning o	avnariment t	ool used to	track Machin	no learning s	secate cuch ac
Neptune.ai is a machi datasets, parameters The assets are tracket	etc. The tool	can be used	d to track, re	trieve and q	uery the ass	ets from each run.

where you're able to view the runs that have been done and view the results of each run.

b7/51

Tutorial: Getting started with Neptune

You have already received a link to a brief Neptune tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with the key features of Neptune that are relevant to this experiment.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:

Experiment's Task

Follow the instructions provided in the experiment task document for the Group-C Neptune phase.

Link to the instructions:

Follow the link for the python:

Questions (Neptune):

Retrieving & Querying Tracked Data:

You will be asked to answer a number of retrieving and querying related questions. You are expected to use the Neptune tool alone to assist you in answering the questions (Do not consult the task instruction document). Try your best to provide the correct answer.

Which of the 13 runs performed best? (i.e. which has the lowest RMSE score?)

Run 7

What is the RMSE value for that run?

0.512

Which of the algorithms (LR & RFR) performed best (lowest RMSE) in their first run?
○ LR
O I don't know
What data features were used for the experimental run with the highest r2_score?
Empty
Compare Run 4 and Run 1, which one had the highest mean absolute error?
Run 4
Run 1
O I don't know
What was the value?
0.602

value for that run.

EX-5: 0.3, EX-11: 0.3

Compare "EX-5" ,"E	X-7", "EX-9" and "	EX-13" in the dashb	ooard table.	
	EX-5	EX-7	EX-9	EX-11
Highest RMSE		\circ	\circ	\circ
Highest R2	\circ	•	\circ	0
Highest mean absolute error		0	0	0
If we want to reproc model was used for		f previous runs, we	need to retrieve	the model. Which
RandomForestRe	gressor(max_depth	=5, min_samples_spli	t=6, n_estimators=5	0)
LinearRegression	(normalize=False)			
RandomForestRe	gressor(max_depth	=10, min_samples_sp	lit=12, n_estimators	=45)
Other:				
From the dashboard	d table filter out	the to show the tra	acked experimen	t runs:
Query Neptune for toparameters used.	the model with th	ne worst RMSE (Larg	gest value). Provic	de Run id, and the
EX-5, normalize=False				
List all linear regress	sion runs with RM	SF value greater th	an 0.9. Provide Ru	un id and the r2

In the python file, print the R2 value of the very first run. What was the value? You are expected to retrieve this information via the python script.						
0.50123						
Experience with No	eptune:					
How helpful was th	he brief Nep	otune tutor	ial provide	d ahead of	this experi	ment?
	1	2	3	4	5	
Not Helpful	0	0	0	0	•	Very Helpful
How do you rate th	he ease of c	completing	the tasks?			
	1	2	3	4	5	
Not Easy	0	0	0	0	•	Very Easy
How helpful is the visual dashboard provided by Neptune.ai when comparing the experimental runs?						
	1	2	3	4	5	
Not Helpful	0	0	0	0	•	Very Helpful

How long did it take you to complete the Neptune.ai section?
O - 30 mins
30 mins - 1 hour
1 hour - 1 hour 30 mins
1 hour 30 mins - 2 hours
2+ hours
Any Additional Comments:

Experience with all Tools (DVC, No-Tool, and Neptune)

Even though the experiment has guided you through few runs of machine learning experiments, a typical machine learning task usually have many more runs carried out over a longer period of time. Consequently, it becomes harder for users to manage the multiple experiment runs and their assets without any supporting tools.

To answer the following questions, we want you to keep a perspective of a machine learning expert who is probably working on multiple ML tasks with way too many experimental runs.

Describe your experience with each of the tools(Neptune, DVC, and No-Tool).

Dvc was my favorite because it's run through the CLI, Neptune dashboard was good but the python file was too messy, no-tool was difficult to track

The DVC and Neptune tools provide a significant support for tracking, querying and retrieving generated data from ML experiments over No-tool.
Strongly Agree
Agree
O Neutral
Disagree
Strongly Disagree
Please elaborate on your response above. When using the no-tool, I found it hard to complete the tasks, whereas neptune and dvc helped me look for the answers
Which tool do you consider best for tracking data during ML experiments ?
Neptune.ai
No-tool
Which tool do you consider best for querying and retrieving previously tracked data?
Neptune.ai
O DVC
O No-tool

Which of Neptune and DVC do you consider least intrusive in completing the tasks?
Neptune.ai
● DVC
Which of Neptune and DVC was the easiest to learn?
Neptune.ai
● DVC
Which of the tools provides the best support for comparing different experiment runs?
Neptune.ai
O DVC
Neptune helps compare different runs using a web dashboard, while DVC uses CLI. Which do you most convenient?
Neptune (Web dashboards)
DVC (CLI)
Please elaborate.
I use git a lot, so using dvc was not that different

Finally, which tool would you recommend a ML practitioner to use?
O Neptune.ai
DVC
○ No-Tool
Please elaborate.
Because it would take long to learn it, especially if you have a git background
Group C:
The following tasks are designed for participants assigned to group C only. You will be guided through three different phases in the following order:

* Phase 1 - No-Tool * Phase 2 - Neptune.ai

* Phase 3 - DVC

Phase 1: No Tool

In this part of the experiment, you will run the experiment with python only. And will you will not have any experiment tools to assist you.

Tutorial: Getting started with No-Tool

You have already received a link to a brief No-Tool tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with running the project with No-Tool.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:

Follow the instructions provided in the experiment task document for the Group-A Neptune phase.
Link to the instructions:
Link to python script:
Questions (No-Tool):
Retrieving & Querying data: You will asked to perform a number of retrieving tasks, and you will use the tool to assist you in performing these tasks. Try your best to provide the correct answer:
Which run performed best, (i.e. has the lowest RMSE score)?
What is the RMSE value for that run?
Which of the algorithms performed best in their first run? LR or RFR i.e. which one had the lowest RMSE.
○ LR
RFR
O I don't know
What data features were used for the run with the lowest r2_score?

Between Run 4 and Run 1, which one had the highest mean absolute error?								
Run 4								
Run 1								
O I don't know								
And what was the value?								
Compare "Run 5", "Run 7", "Run 9" and "Run 13":								
	Run 5	Run 7	Run 9	Run 13				
Highest RMSE	0	0	\circ	\circ				
Highest R2	\circ	\circ	\circ	\circ				
Highest mean absolute error								
If we want to reproduce the results of previous runs, we need to retrieve the model. Which model was used for "Run 4"?								
RandomForestRe	gressor(max_depth=	5, min_samples_spl	it=6, n_estimators=5	0)				
LinearRegression	0							
RandomForestRe	gressor(max_depth=	:10, min_samples_sp	olit=12, n_estimators	=45)				
Other:								

For the following write down the run number along with the value:									
Run with the worst RMSE (Largest value) Provide the run nr and RMSE									
Where you used linear regression algorithm and the RMSE was less than 0.7 Provide Run number, and the r2 value for that run.									
What was the R2 value of the very first run?									
Experience with No-Tool:									
How helpful was t	How helpful was the short tutorial in the beginning in completing the tasks?								
	1	2	3	4	5				
Not Helpful	0	0	0	0	0	Very Helpful			
How do you rate the ease of completing the tasks?									
	1	2	3	4	5				
Not Easy	0	0	0	0	0	Very Easy			

low helpful was yo uns and their asso		a manual te	chnique in	tracking/re	etrieving th	e experimental
	1	2	3	4	5	
Not Helpful	0	0	0	0	0	Very Helpful
How long did it tak	e you to co	mplete the	e No-Tool s	ection?		
0-30 mins						
30 mins-1 hour						
1 hour-1 hour 30) mins					
1 11001-1 11001/30						
1 hour 30mins-2	2 hours					
	? hours					

Data Version Control is an experiment management tool that extends Git to make ML models shareable and reproducible. It is designed to handle large files, datasets, machine learning models, metrics as well as code.

Tutorial: Getting started with DVC

You have already received a link to a brief DVC tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with the key features of DVC that are relevant to this experiment.

We have also provided the link below in case you would like to review it.

Follow the link for the tutorial:



Experiment's Task

Follow the instructions provided in the experiment task document for the Group-C DVC phase.

Link to the instructions:

Questions (DVC):

Retrieving & Querying data:

You will asked to perform a number of retrieving tasks, and you will use the tool to assist you in performing these tasks. Try your best to provide the correct answer:

Which run performed best, (i.e. has the lowest RMSE score)?
What is the RMSE value for that run?

Which of the algorithms performed best in their first run? LR or RFR i.e. which one had the lowest RMSE.
○ LR
RFR
O I don't know
What data features were used for the run with the highest r2_score?
Between Run 4 and Run 1, which one had the highest mean absolute error?
Run 4
Run 1
O I don't know
What was the value?

Compare "Run 5", "Run 7", "Run 9" and "Run 13":											
	Run 5	Run 7	Run 9	Run 13							
Highest RMSE											
Highest R2	Highest R2										
Highest mean absolute error											
If we want to reproduce the results of previous runs, we need to retrieve the model. Which model was used for "Run 4"?											
RandomForestReg	RandomForestRegressor(max_depth=5, min_samples_split=6, n_estimators=50)										
LinearRegression()											
RandomForestRegressor(max_depth=10, min_samples_split=12, n_estimators=45)											
Other:											
Using DVC commands, solve the following:											
Return the run with the worst RMSE (Largest value) Provide tag name, and the parameters used.											
Where you used linear regression algorithm and the RMSE was less than 5 Provide tag name, and the parameters used.											

What was the R2 value of the very first run? Hint: Use dvc metrics show, as shown in tutorial									
Experience with DVC:									
How helpful was the short tutorial in the beginning in completing the tasks?									
1 2 3 4 5									
Not Helpful	0	0	0	0	0	Very Helpful			
How do you rate the ease of completing the tasks?									
	1	2	3	4	5				
Not Easy	0	0	0	0	0	Very Easy			
How helpful were the DVC commands to compare the runs?									
	1	2	3	4	5				
Not Helpful	0	0	0	0	\circ	Very Helpful			

How long did it take you to complete the DVC section?
O - 30 mins
30 mins - 1 hour
1 hour - 1 hour 30 mins
1 hour 30 mins - 2 hours
2+ hours
Any Additional Comments:
Phase 3: Neptune
Neptune.ai is a machine learning experiment tool used to track Machine learning assets such as datasets, parameters etc. The tool can be used to track, retrieve and query the assets from each run. The assets are tracked as metadata which can be viewed in the GUI. It provides the user with a GUI where you're able to view the runs that have been done and view the results of each run.
Tutorial: Getting started with Neptune You have already received a link to a brief Neptune tutorial that we have prepared for this experiment. We expect that you have familiarized yourself with the key features of Neptune that are relevant to this experiment. We have also provided the link below in case you would like to review it. Follow the link for the tutorial:

Experiment's Task

Follow the instructions provided in the experiment task document for the Group-C Neptune phase.

Link to the instructions:

Follow the link for the python:

Questions (Neptune):

Neti levilla & Quei villa i lackea Dat	ing Tracked Data:	Query	&	ievina	Retri	F
--	-------------------	-------	---	--------	-------	---

You will be asked to answer a number of retrieving and querying related questions. You are expected to use the Neptune tool alone to assist you in answering the questions (Do not consult the task instruction document). Try your best to provide the correct answer.

Which of the 13 runs performed best? (i.e. which has the lowest RMSE score?)
What is the RMSE value for that run?
Which of the algorithms (LR & RFR) performed best (lowest RMSE) in their first run? LR
RFR I don't know
What data features were used for the experimental run with the highest r2_score?

Compare Run 4 and Run 1, which one had the highest mean absolute error?								
Run 4								
Run 1								
O I don't know								
What was the value?								
Compare "EX-5", "EX-7", "EX-9" and "EX-13" in the dashboard table.								
	EX-5	EX-7	EX-9	EX-11				
Highest RMSE	0	\circ	\circ	\bigcirc				
Highest R2	0	\circ	\circ	0				
Highest mean absolute error	0	0						
If we want to reproduce the results of previous runs, we need to retrieve the model. Which model was used for "EX-4"?								
RandomForestRe	gressor(max_depth:	=5, min_samples_spl	it=6, n_estimators=5	0)				
LinearRegression	()							
RandomForestRe	gressor(max_depth:	=10, min_samples_sp	olit=12, n_estimators	=45)				
Other:								

From the dashboa	rd table filt	ter out the	to show t	he tracked	l experime	nt runs:	
Query Neptune for parameters used.	the model	with the w	vorst RMSE	(Largest v	alue). Prov	de Run id, and the	
List all linear regres	ssion runs v	vith RMSE	value less t	han 54. Pro	ovide Run i	d, and the r2 value	
In the python file, print the R2 value of the very first run. What was the value? You are expected to retrieve this information via the python script.							
Experience with Ne	eptune:						
How helpful was th	e brief Nep	otune tutor	ial provide	d ahead of	this experi	ment?	
	1	2	3	4	5		
Not Helpful	0	0	0	0	0	Very Helpful	
How do you rate th	e ease of c	completing	the tasks?	,			
	1	2	3	4	5		

How do you rate the ease of completing the tasks?

1 2 3 4 5

Not Easy

Very Easy

How helpful is the very experimental runs?		board prov	vided by Ne	eptune.ai w	hen compa	aring the
	1	2	3	4	5	
Not Helpful	0	0	0	0	0	Very Helpful
How long did it take	e you to co	mplete the	e <u>neptune.a</u>	<u>ai</u> section?		
0 - 30 mins						
30 mins - 1 hour						
1 hour - 1 hour 3	0 mins					
1 hour 30 mins -	2 hours					
2+ hours						
Any Additional Con	nments:					
Experience with all	Tools (No-	Tool, DVC,	Neptune)			
Even though the experitypical machine learning Consequently, it becon without any supporting	ng task usua nes harder f	ally have ma	ny more run	s carried ou	t over a long	er period of time.
To answer the followin is probably working on						e learning expert who
Describe your expe	erience witl	h each of t	he tools(Ne	eptune, DV	C and No-1	Tool).

The DVC and Neptune tools provide a significant support for tracking, querying and retrieving generated data from ML experiments over No-tool.					
Strongly Agree					
Agree					
Neutral					
O Disagree					
Strongly Disagree					
Please elaborate on your response above.					
Which tool do you consider best for tracking data during ML experiments?					
Neptune.ai					
O DVC					
No-tool No-tool					
Which tool do you consider best for querying and retrieving previously tracked data?					
Neptune.ai					
O DVC					
No-tool					

Which of Neptune and DVC do you consider least intrusive in completing the tasks?				
Neptune.ai				
O DVC				
Which of Neptune and DVC was the easiest to learn?				
Neptune.ai				
O DVC				
Which of the tools provides the best support for comparing different experiment runs?				
Neptune.ai				
O DVC				
Neptune helps compare different runs using a web dashboard, while DVC uses CLI. Which do you most convenient?				
Neptune (Web dashboards)				
O DVC (CLI)				
Please elaborate.				

Finally, which tool would you recommend a ML practitioner to use?
Neptune.ai
O DVC
O No-Tool
Please elaborate.
Conclusion
Thank you for participating.

This content is neither created nor endorsed by Google.

Google Forms