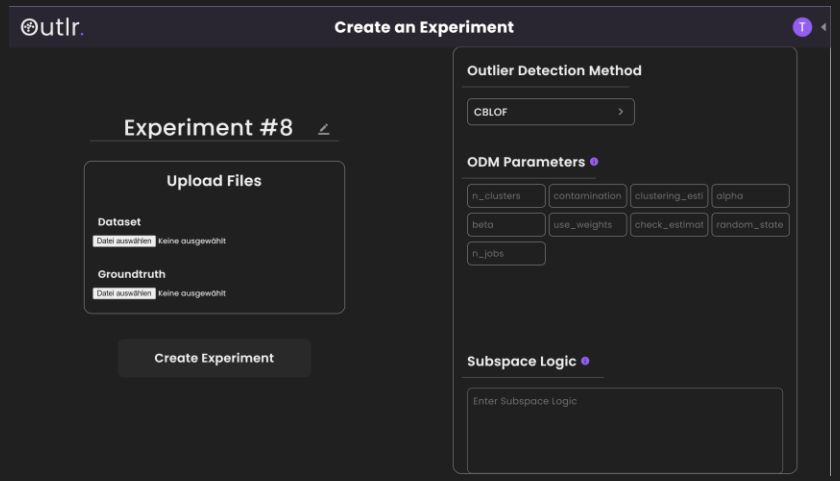
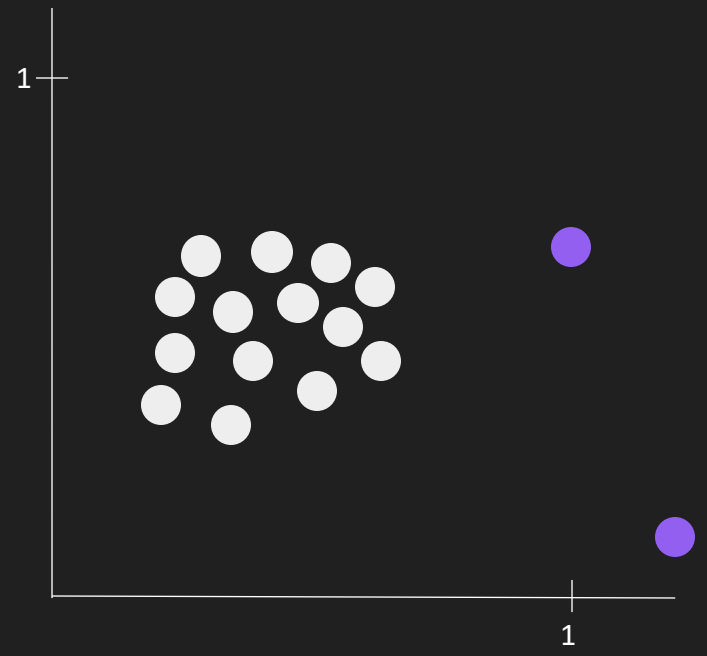
Bennet Hörmann, Salomo Hummel, Simeon Schrape, Erik Wu, Udo Zucker

Motivation



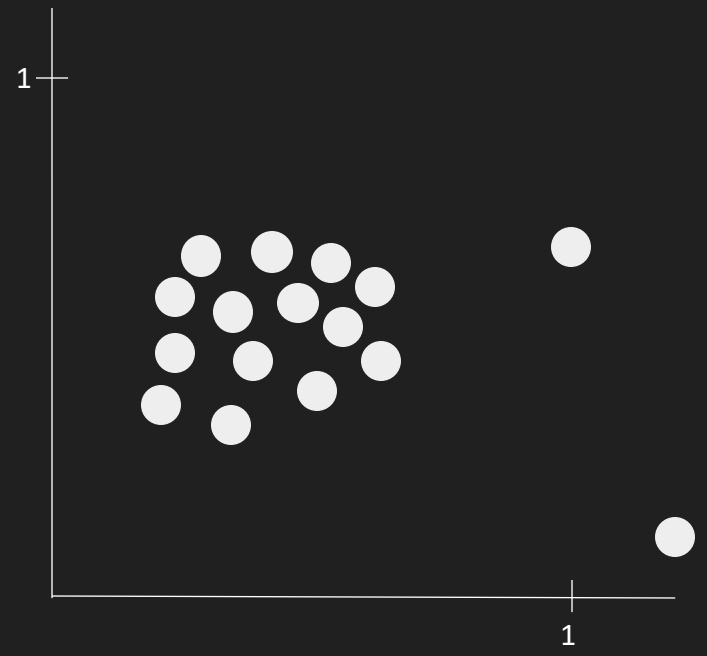
- outliers can degrade ML methods
- not many platforms for subspace outliers
- Outlr. provides accessible + intuitive Interface

Outlier



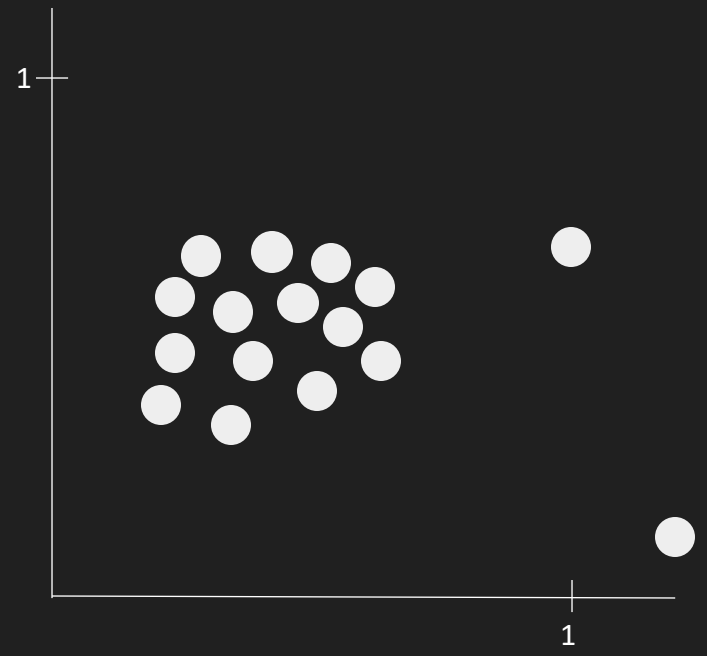
data point that differs significantly from other data points

Subspace Outliers



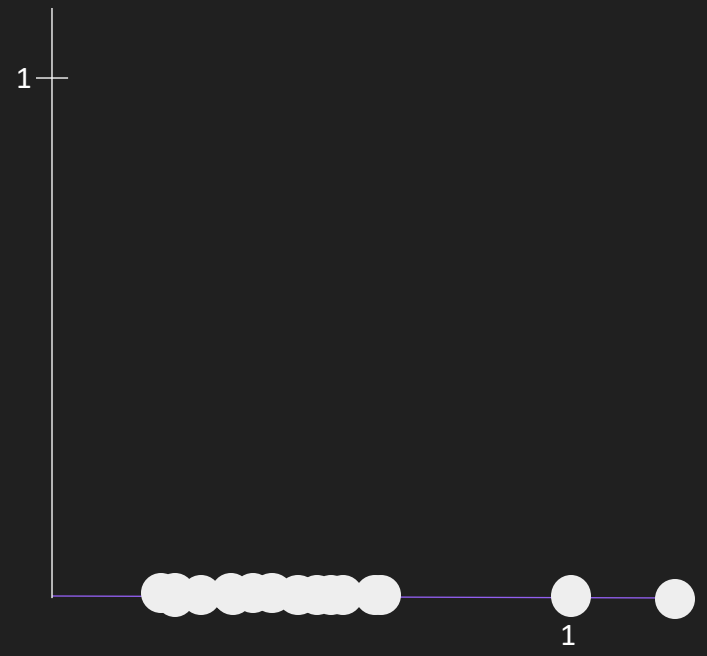
data points which are outliers in certain subspaces

Subspace Outliers



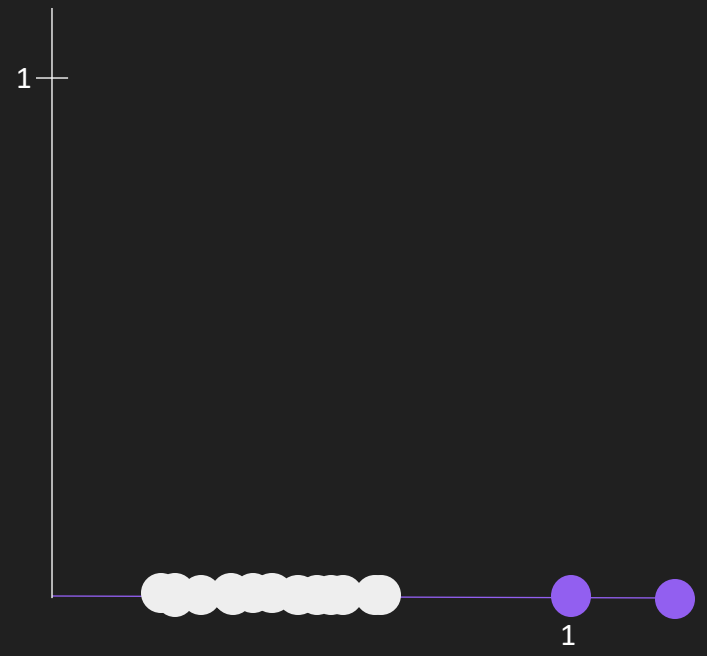
x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Subspace Outliers



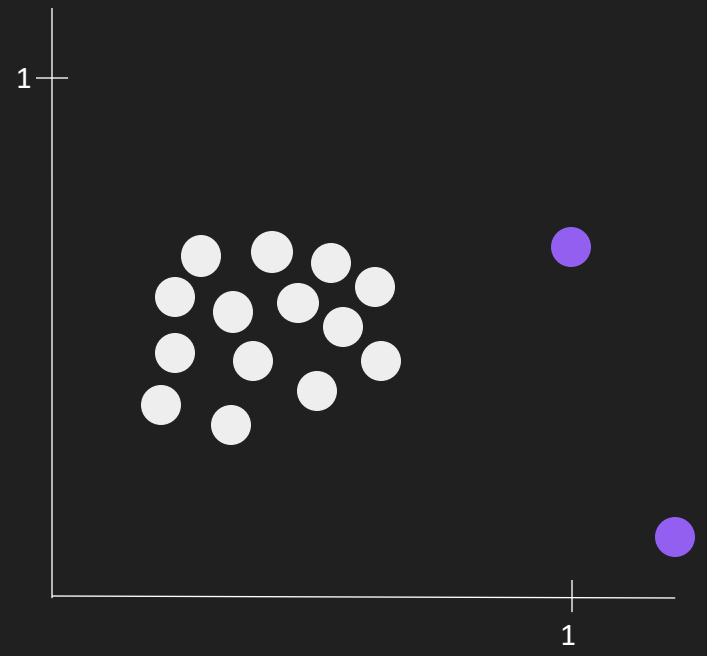
x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Subspace Outliers



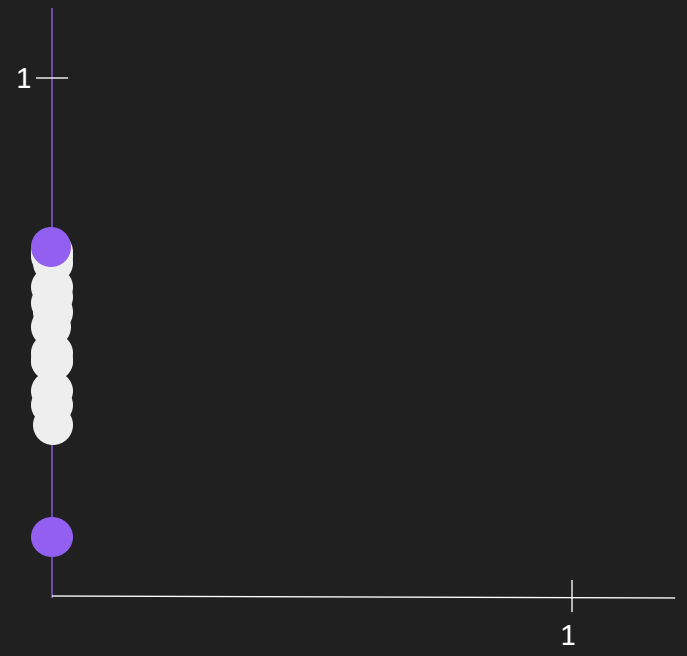
x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Subspace Outliers



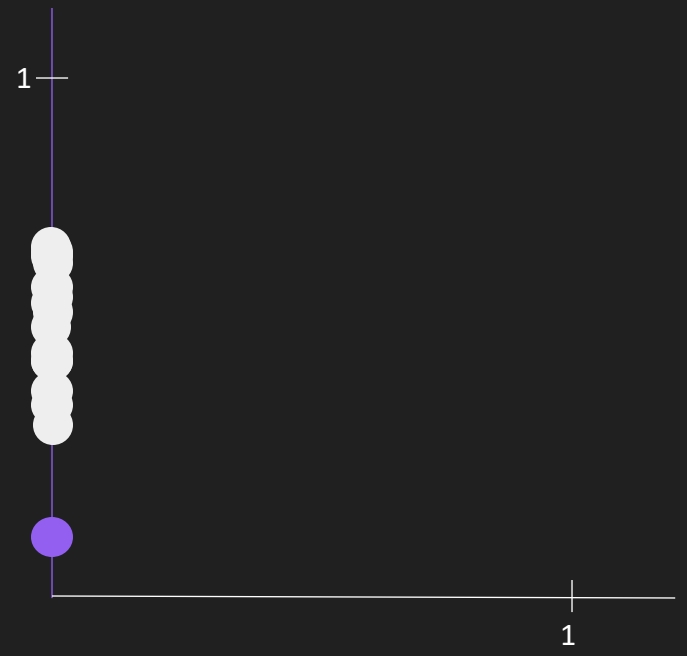
x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Subspace Outliers



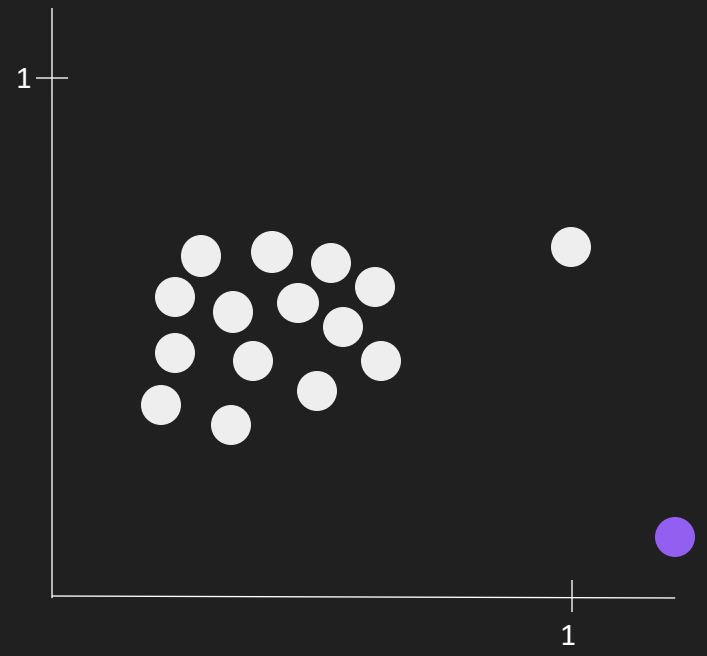
x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Subspace Outliers



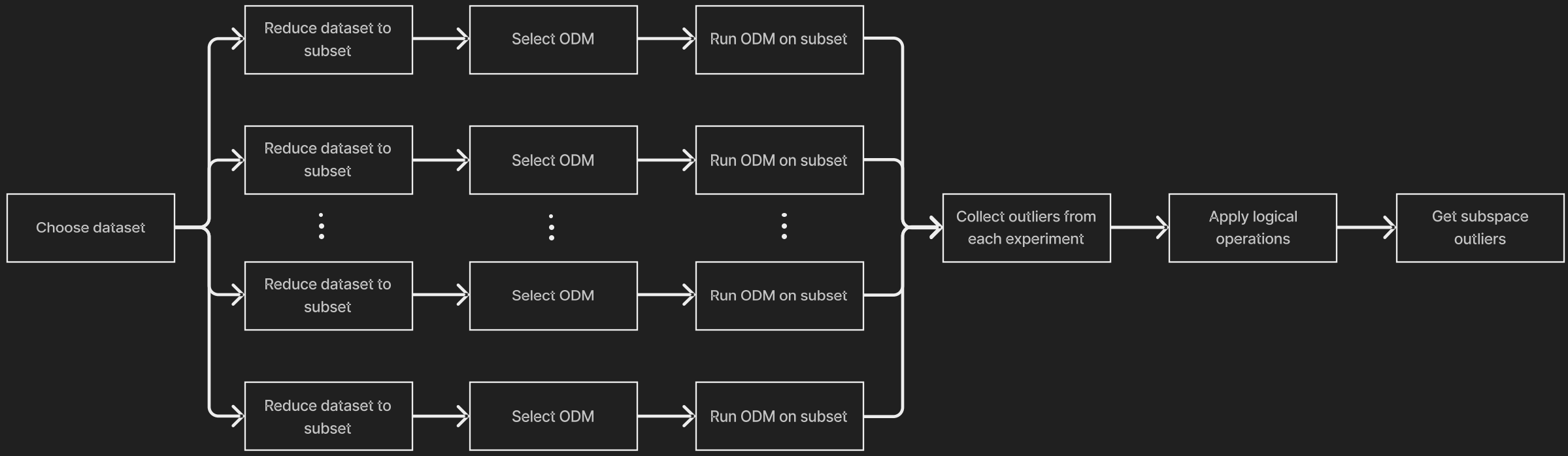
x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Subspace Outliers

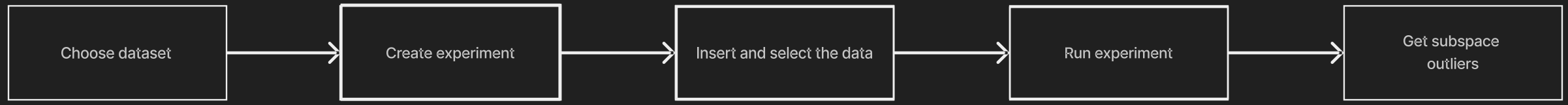


x = „feature 1“	y = „feature 2“
2.8391146111970698	3.1390820095218572
2.8428654177443216	3.1894471104030235
3.305442968608614	2.8697267596737555
3.0054629889589455	3.1304537686350637
2.785299474561798	3.001912971003542
3.1416954644354615	3.2134305096501703
3.1416954644354615	2.935325810188675
1.1204982390	5.432089437928

Problem


Choose dataset
Reduce dataset to subset
Select ODM
Run ODM on subset
Reduce dataset to subset
Select ODM
Run ODM on subset
Reduce dataset to subset
Select ODM
Run ODM on subset
Reduce dataset to subset
Select ODM
Run ODM on subset
Collect outliers from each experiment
Apply logical operations
Get subspace outliers

Solution

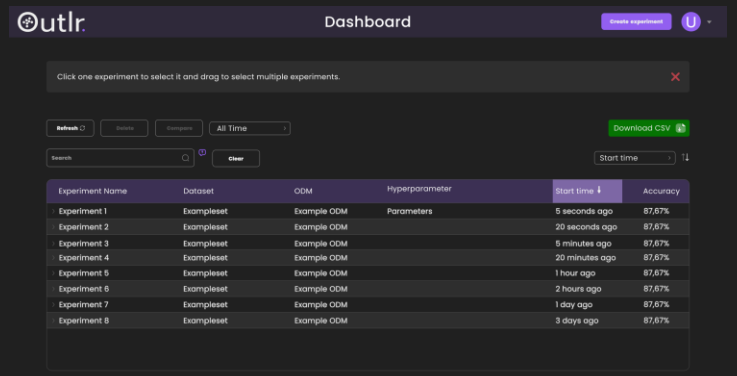

Choose dataset
Create experiment
Insert and select the data
Run experiment
Get subspace outliers

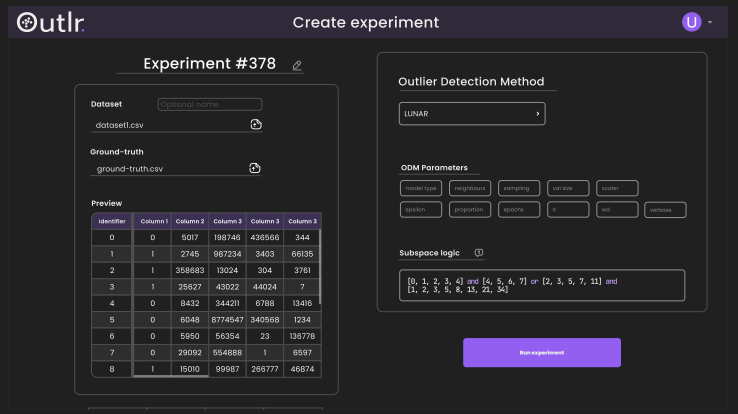
Live Demo

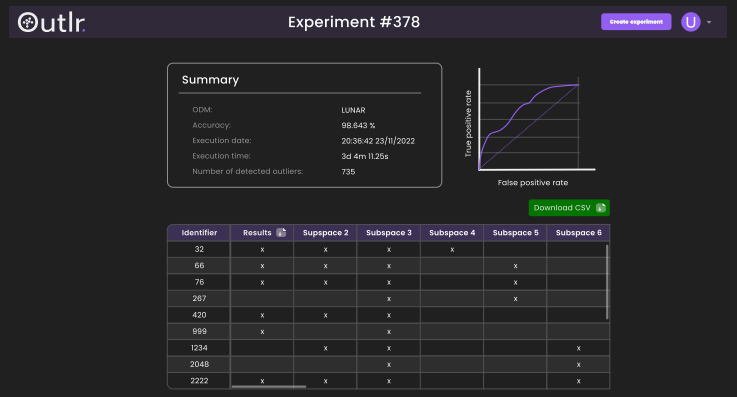
Requirements Specification

Product functions and requirements

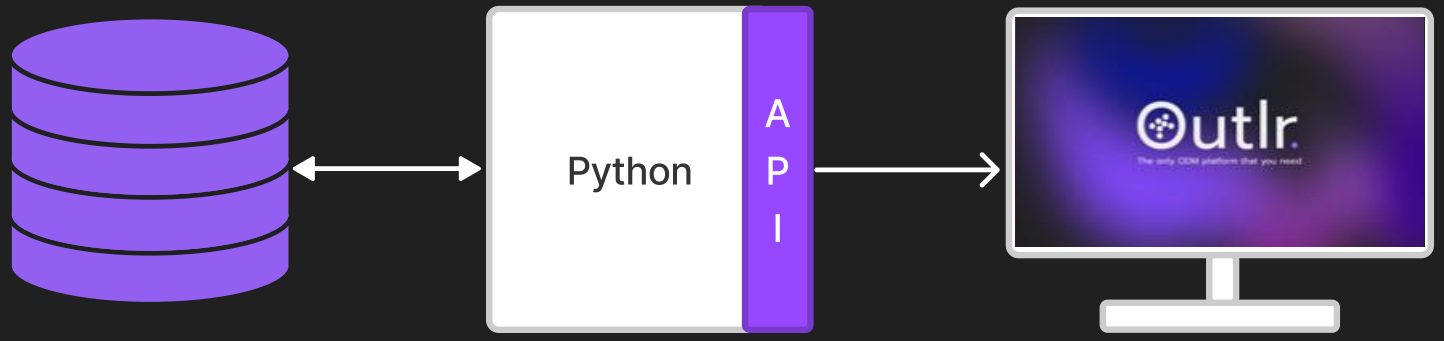
GUI Mockups

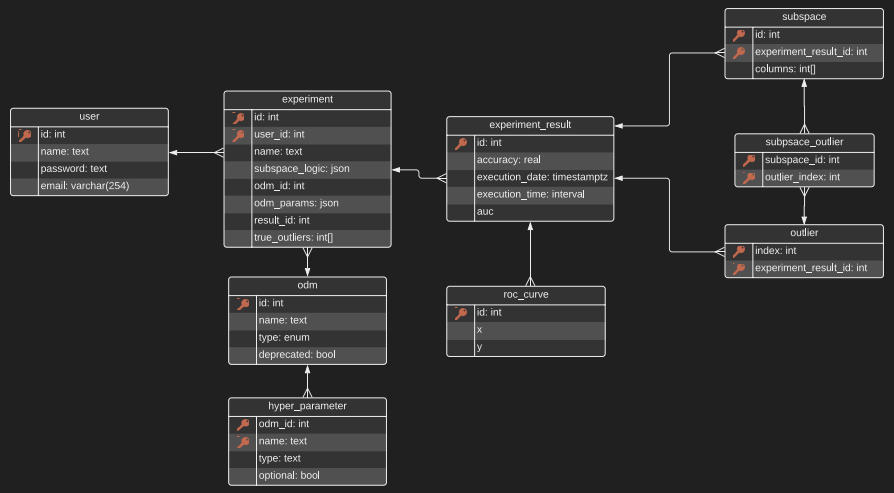


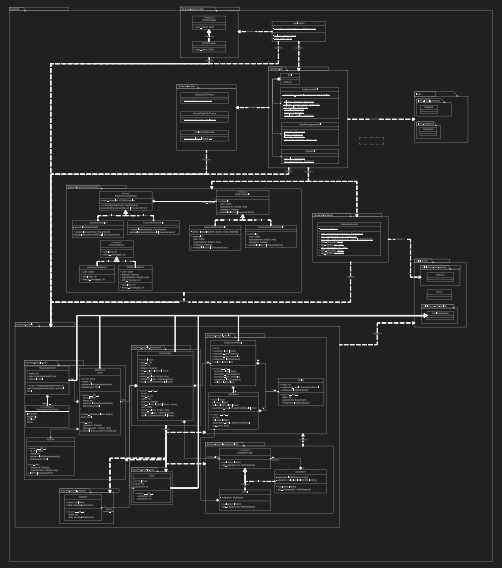


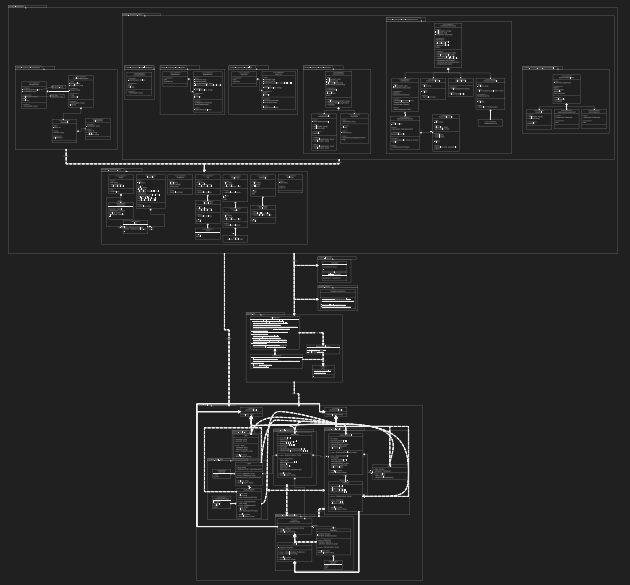


Design


Python
A
P
I
Outlr
SQL Database
Backend
Web Frontend


user
id: int
name: text
password: text
email: varchar(254)
experiment
id: int
user_id: int
name: text
subspace_logic: json
odm_id: int
odm_params: json
result_id: int
true_outliers: int[]
experiment_result
id: int
accuracy: real
execution_date: timestamptz
execution_time: interval
auc
subspace
id: int
experiment_result_id: int
columns: int[]
subpsace_outlier
subspace_id: int
outlier_index: int
outlier
index: int
experiment_result_id: int
odm
id: int
name: text
type: enum
deprecated: bool
roc_curve
id: int
x
y
hyper_parameter
odm_id: int
name: text
type: text
optional: bool





Implementation

26 Vue Components
34 TS Files
3671 Source Code Lines
409 Comment Lines

34 Files
1213 Source Code Lines
608 Comment Lines

Total 4884 Source Code Lines

without empty lines, without test files

Quality Assurance

Frontend **Jest**

90% statement coverage | 78 test cases

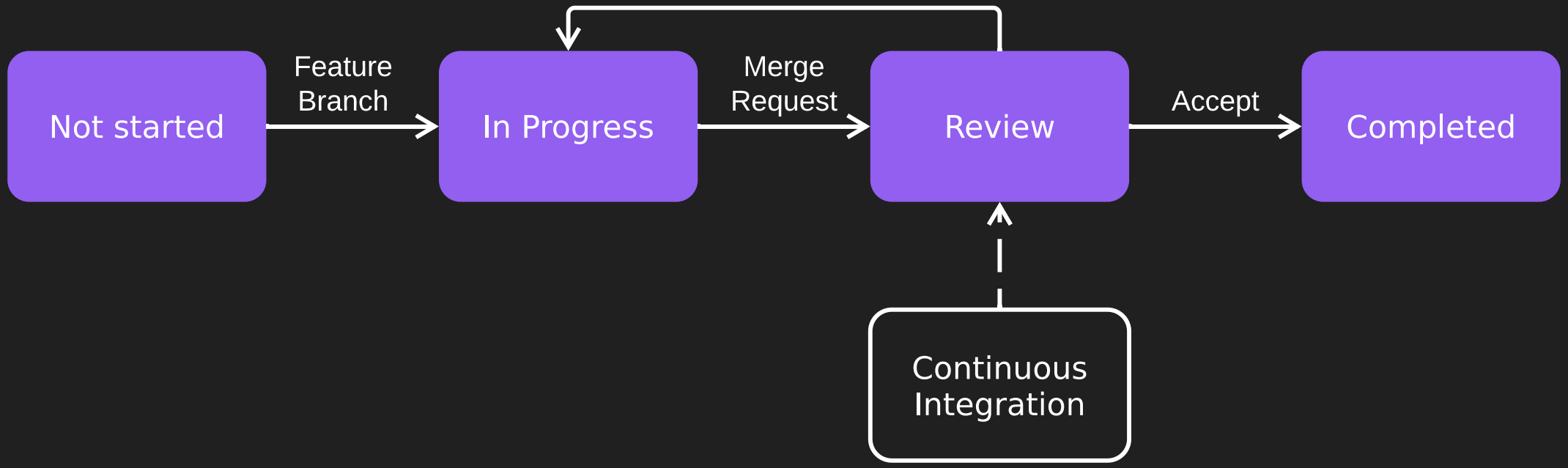
Backend **unittest**

94% statement coverage | 44 test cases

ESLint

flake8

Workflow


Not started
Feature Branch
In Progress
Merge Request
Review
Accept
Completed
Continuous Integration

Difficulties

Communication

Time management

Estimating how long tasks take

Getting used to new tools

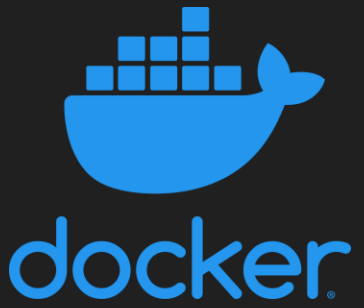
Learned a Lot of Things











Soft Skills

Communication

Team work

Task distribution

Presenting

Logo Sources

- Figma https://de.wikipedia.org/wiki/Figma
- Overleaf https://en.wikipedia.org/wiki/Overleaf
- LaTex https://de.wikipedia.org/wiki/LaTeX
- Git https://de.wikipedia.org/wiki/Git
- GitLab https://de.wikipedia.org/wiki/GitLab
- PyCharm https://de.wikipedia.org/wiki/PyCharm
- Docker https://www.prnewswire.com/news-releases/docker-and-tigera-partner-to-integrate-networking-support-for-kubernetes-on-windows-in-docker-enterprise-300853617.html
- Vue https://en.wikipedia.org/wiki/Vue.js#History
- TypeScript https://en.wikipedia.org/wiki/TypeScript
- Python https://en.wikipedia.org/wiki/Python_(programming_language)
- SQLAlchemy https://www.sqlalchemy.org/
- Flask https://flask.palletsprojects.com/en/2.2.x/
- ESLint https://eslint.org/branding/

Thank you for listening!

Bennet Hörmann, Salomo Hummel, Simeon Schrape, Erik Wu, Udo Zucker