

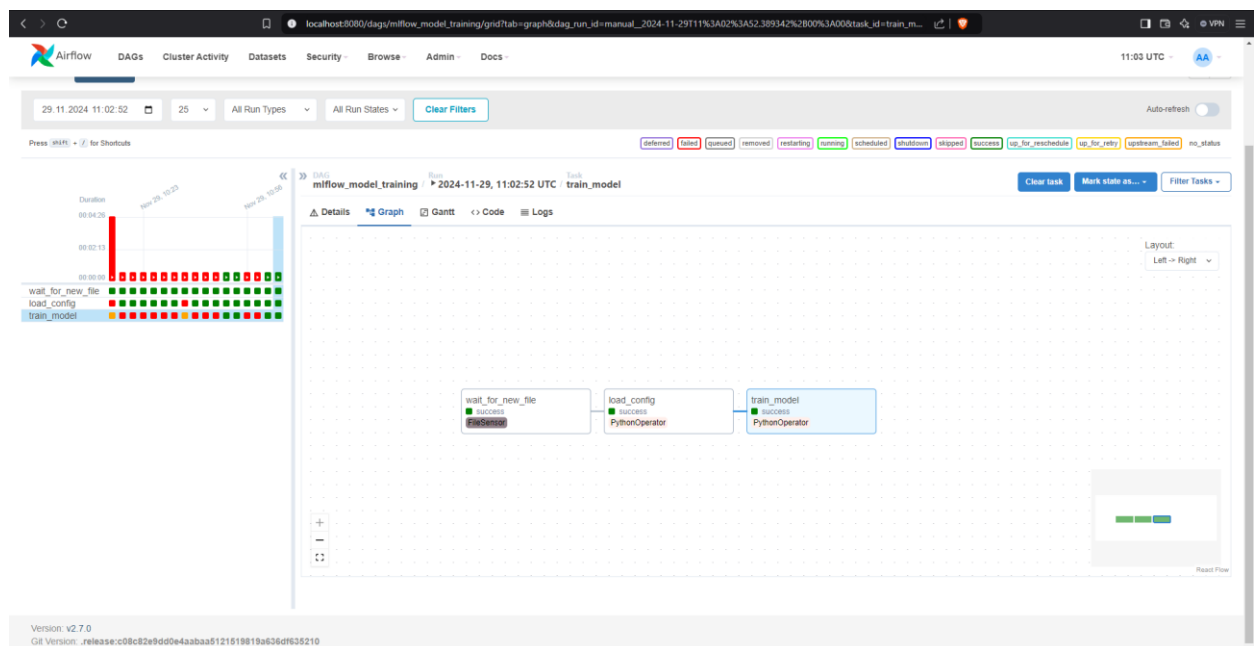
Отчет по лабораторной работе №3

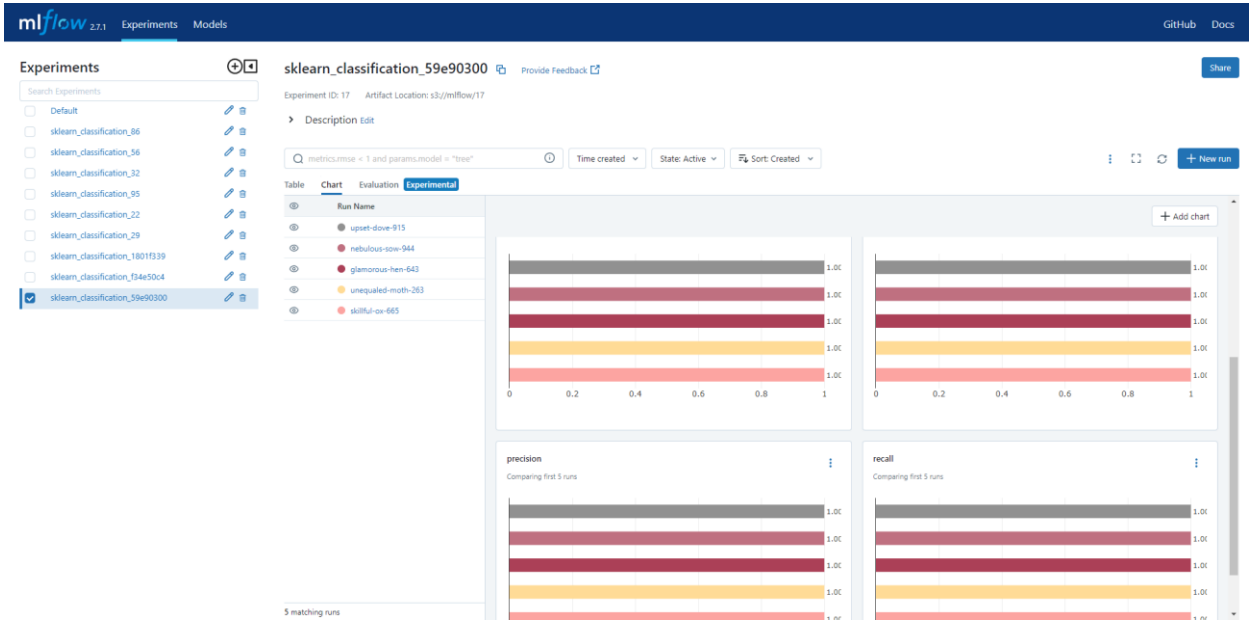
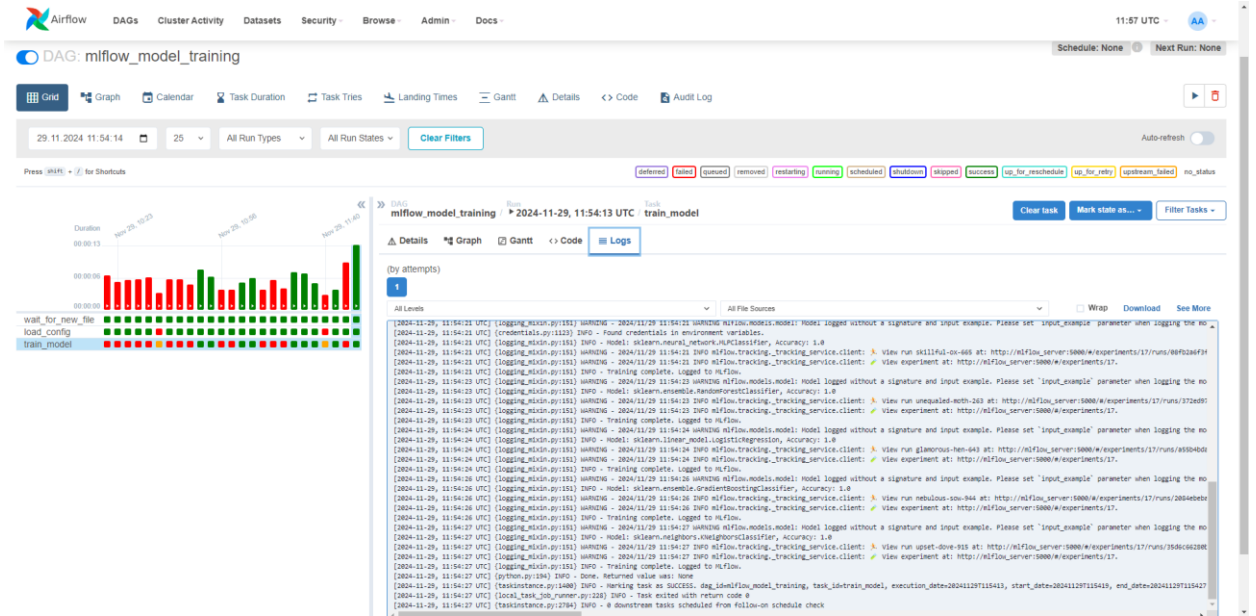
Выполнил: Казаркин Тимофей, гр. 6232

Apache Airflow, MLFlow, обучение модели и логирование

Ход выполнения первого задания:

- По ходу задания было написано несколько заданий для DAG, первое ожидает появления файлов в директории, второе считывает конфигурационные файлы в формате json, третье запускает обучение классификаторов и логирует параметры в mlflow;
- Из интересного стоит отметить, что было необходимо динамически подгружать модули по имени и динамически инициализировать классы классификаторов.





Parameters (6)

Name	Value
algorithm	auto
classifier	sklearn.neighbors.KNeighborsClassifier
leaf_size	30
n_neighbors	5
p	2
weights	uniform

Metrics (4)

Name	Value
accuracy 	1
f1 	1
precision 	1
recall 	1

Name	Value
C	1.0
classifier	sklearn.linear_model.LogisticRegression
max_iter	100
multi_class	auto
penalty	l2
solver	lbfgs

Metrics (4)

Name	Value
accuracy 	1
f1 	1
precision 	1
recall 	1

Name	Value
classifier	sklearn.ensemble.GradientBoostingClassifier
learning_rate	0.1
loss	log_loss
max_depth	3
min_samples_split	2
n_estimators	150

Metrics (4)

Name	Value
accuracy 	1
f1 	1
precision 	1
recall 	1

Name	Value
classifier	sklearn.ensemble.RandomForestClassifier
criterion	gini
max_depth	10
min_samples_leaf	1
min_samples_split	2
n_estimators	100

Metrics (4)

Name	Value
accuracy 	1
f1 	1
precision 	1
recall 	1

Name	Value
activation	relu
classifier	sklearn.neural_network.MLPClassifier
hidden_layer_sizes	[20, 100, 20]
learning_rate	constant
learning_rate_init	0.001
solver	adam

▼ Metrics (4)

Name	Value
accuracy 	1
f1 	1
precision 	1
recall 	1

mlflow2.7.1

Experiments

Models

GitHubDocs

Registered Models

Create Model

Filter registered models by name or tags

Name ⓘ	Latest version	Staging	Production	Created by	Last modified	Tags
sklearn.ensemble.GradientBoostingClassifier	Version 1	—	—		2024-11-29 16:12:59	—
sklearn.ensemble.RandomForestClassifier	Version 1	—	—		2024-11-29 16:12:53	—
sklearn.linear_model.LogisticRegression	Version 1	—	—		2024-11-29 16:12:56	—
sklearn.neighbors.KNeighborsClassifier	Version 1	—	—		2024-11-29 16:13:01	—
sklearn.neural_network.MLPClassifier	Version 1	—	—		2024-11-29 16:12:50	—

Apache Airflow, MLFlow, перевод модели в стадию production

Ход выполнения второго задания:

- Было написано задание для DAG, которое выбирает лучшую модель и переводит её в стадию production;
- Из интересного хотелось бы отметить, что в apache airflow нет поддержки библиотеки huggingface_hub, и я не успел разобраться с её установкой, так что выполнить хост модели не вышло, хотя код был написан;
- Лучшая модель выбиралась по наивысшему значению accuracy, хотя с заданными параметрами все классификаторы выдали значения 1.0 accuracy.

mlflow2.7.1

ExperimentsModels

Registered Models

Filter registered models by name or tags

Name ↕	Latest version	Staging	Production	Created by	Last modified	Tags
sklearn.ensemble.GradientBoostingClassifier	Version 1	—	Version 1		2024-11-30 22:29:21	—
sklearn.ensemble.RandomForestClassifier	Version 1	—	—		2024-11-29 16:12:53	—
sklearn.linear_model.LogisticRegression	Version 1	—	—		2024-11-29 16:12:56	—
sklearn.neighbors.KNeighborsClassifier	Version 1	—	—		2024-11-29 16:13:01	—
sklearn.neural_network.MLPClassifier	Version 1	—	—		2024-11-29 16:12:50	—