

# Frameworks zum Management von Data Science-Experimenten

Wesam Al-Shaibani

Bachelorarbeit • Studiengang Informatik • Fachbereich Informatik und Medien • 26.01.2022

## Aufgabenstellung

Das Ziel dieser Bachelorarbeit liegt in der Gegenüberstellung von den zwei Frameworks Neptune.ai und Comet.ml, die zum Management von Data Science-Experimenten dienen.

## Experimentmanagementsysteme

Projekte des maschinellen Lernens beinhalten das mehrmalige Ausführen von Modellen, damit diese Modelle bei der Produktion übernommen werden. Es werden unterschiedliche Datenversionen, Quellcodes, Hyperparameter und Umgebungs-konfigurationen und noch mehr verwaltet. Die Komplexität des Projekts erhöht sich noch weiter, wenn weitere Datenobjekte wie Artefakten dazu mitgezählt werden. Die Entstehung von Modellen in der Produktion sollte ebenfalls nachvollziehbar sein [1]. Zu diesem Zweck stehen Experimentmanagementsysteme zur Verfügung, die Experimente verfolgen und reproduzieren können sollen.

## Konzept und Implementierung

Mithilfe eines Konvolutionellen Neuronalen Netzwerks (CNN) und eines Mehrschichtigen Perzeptrons (MLP) werden Bildklassifikationsmodelle für den MNIST-Datensatz erstellt. Die Experimente werden in den zwei Frameworks Neptune.ai und Comet.ml protokolliert und von denen verfolgt. In den Experimenten werden die Unterschiede der beiden Frameworks anhand bestimmter Kriterien untersucht. Diese Kriterien beinhalten die Umgebung, die Quellcodes, die Datensätze, die Modelle, die Bilder, die Artefakten, die Benutzeroberfläche, die Hyperparameter und Metriken. Die Visualisierungen zu den Testergebnissen der Bildklassifikationsmodellen werden automatisch von Neptune.ai und Comet.ai erstellt. Zwei Conda-Umgebungen mit den installierten Python-Paketen für das jeweilige Framework, wurden auch erstellt. Die Experimente wurden lokal in einem Jupyter Notebook ausgeführt und für die Versionskontrolle wurde Git verwendet.

## Ergebnisse

Die Ergebnisse können der Tabelle 1 entnommen werden.

Neptune.ai	Comet.ml
Modellregistrierung und Modellversionierung sind noch nicht verfügbar, jedoch Modellerfassung ist möglich.	Modellerfassung, Modellregistrierung, und Modellversionierung sind verfügbar.
In Neptune werden Ordner und Pfade (Experimentreiter) meistens selbst erstellt.	Die meisten Ordner und Pfade (Experimentreiter) in Comet.ml werden automatisch mit jedem Experiment erstellt.

Datensätze, Modelle und Bilder können als Artefakten protokolliert werden, jedoch diese sind nur innerhalb eines Experiments verfügbar.	Datensätze, Modelle und Bilder können unter anderem sowohl als Artefakten als auch Assets protokolliert werden. Artefakten sind auch außerhalb eines Experiments verfügbar.
Metriken, Parameter, Quellcodes können bisher in Neptune.ai nur manuell erfasst werden.	Metriken, Parameter, Quellcodes und so weiter können in Comet.ml automatisch erfasst werden.
Datensätze können protokolliert (hochgeladen) werden.	Informationen über die Datensätze können protokolliert werden. Die Datensätze selbst werden z.B. als Artefakten oder Assets protokolliert.

Tabelle 1: Ergebnisse

## Fazit

In dieser Bachelorarbeit wurden zwei Frameworks zum Management von Data Science-Experimenten, Neptune.ai und Comet.ml, verglichen. Aus den Experimenten mit Neptune.ai und Comet.ml wurde festgestellt, dass die beiden Frameworks viele Gemeinsamkeiten haben, jedoch sich in einigen Aspekten unterscheiden. Die Möglichkeit zur automatischen Erfassung von Daten wie Metriken, Parameter, und so weiter in Comet.ml ist ein Grund dafür, Comet.ml vorzuziehen. In Aspekten wie mehr Freiheit bei der Wahl von Ordnern und Pfaden zum Erfassen von Daten wird Neptune.ai bevorzugt.

## Quellen

[1] Eble, Michael. Managing Machine Learning Cycles: Five Learnings from comparing Data Science Experimentation/ Collaboration Tools. KDnuggets. [online] (29.01.2020). <https://www.kdnuggets.com/2020/01/managing-machine-learning-cycles.html>. [Zugriff am: 22.01.2022].

neptune.ai. Neptune: experiment management and collaboration tool. [online](2020). <https://neptune.ai>. [Zugriff am: 21.01.2022]

home page. Comet.ML. <https://www.comet.ml>. [online] [Zugriff: 21.01.2022].