

Entwicklung eines modernen Lehr- und Lernkonzepts für Maschinelles Lernen mit Infrastructure-as-a-Service und Cloud-Computing

Patrick Palsbröker, Matthias Lutz, Carsten Gips und Matthias König¹

Abstract: Im Rahmen des BMBF-geförderten Projektes „Machine Learning Campus Minden“ wurde ein Lehr- und Lernkonzept entwickelt, um den Schwerpunkt Maschinelles Lernen (ML) im Curriculum des Masterstudiengangs Informatik der Fachhochschule Bielefeld auszubauen. Dazu wird das Grundlagenmodul “Methoden des Maschinellen Lernens” zur Vermittlung des theoretischen Basiswissens angeboten. Darauf aufbauend können die Studierenden ihre Kenntnisse in einer Reihe von anwendungsorientierten Vertiefungsmodulen wie beispielsweise “Natural Language Processing” (NLP) oder “Data Science“ (DS) an Fragestellungen aus dem Forschungs- und Industriealltag erproben und erweitern. Um eine praxisnahe Ausbildung auf dem aktuellen technologischen Stand zu gewährleisten, bieten alle neuen Module die praktische Anwendung mittels modernen Cloud-Plattformen an. In diesem Beitrag wird das so entstandene Lehr- und Lernkonzept vorgestellt. Außerdem stellen wir die gesammelten Erfahrungen bei der Integration von Infrastructure-as-a-Service-Diensten in der Lehre dar und gehen sowohl auf die fachlichen als auch organisatorischen Anforderungen und Fallstricke ein sowie auf die Evaluation der eingesetzten Cloud-Anbieter in Hinblick auf die domänenspezifischen und praxisrelevanten Probleme.

Keywords: Lehrkonzept, Lernkonzept, Maschinelles Lernen, Cloud-Computing, Google Cloud Platform, Amazon Web Services, Microsoft Azure.

1 Einleitung

Maschinelles Lernen in seinen vielen Facetten erhält eine immer stärkere Bedeutung, weshalb es uns notwendig erschien, die Lehre in diesem Bereich deutlich auszubauen. Die Verarbeitung von großen Datenmengen erfordert sehr leistungsfähige Hardware, deren Anschaffung und Wartung gerade für Hochschulen eine Herausforderung darstellt [YF15], zumal die für das Maschinelle Lernen benötigte Hardware schnell veraltet. Nach Umfragen aus dem Jahr 2014 stellen lediglich ein Drittel der Hochschulen Rechencluster bereit, so dass es in Lehrveranstaltungen bei einer theoretischen Betrachtung der Methoden bleibt [ST14]. Wir stellen daher in diesem Artikel ein Lehr- und Lernkonzept zur praxisnahen Ausbildung im Bereich Maschinelles Lernen sowie ein Konzept zur flexiblen Integration von Cloud-Diensten im Rahmen der Lehre vor.

¹ FH Bielefeld Campus Minden, Artilleriestraße 9, 32427 Minden {Vorname.Nachname}@fh-bielefeld.de

2 Lehr- und Lernkonzept

Das im Rahmen des vom BMBF geförderten Projektvorhabens entwickelte Lehr- und Lernkonzept setzt sich, wie in Abb. 1 dargestellt, aus drei aufeinander aufbauenden Ebenen zusammen: Basismodul, Vertiefungsmodule und abschließende Masterarbeit.

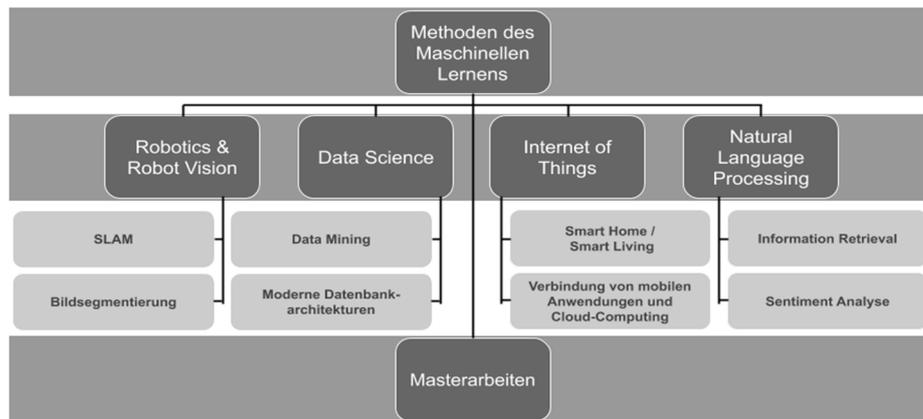


Abb. 1: ML-Camp Lehr- und Lernkonzept

Im Basismodul werden aktuelle Algorithmen und Methoden aus dem Bereich ML umfangreich vermittelt, wobei der Fokus auf Algorithmen des überwachten und unüberwachten Lernens liegt. Ein wichtiger Bestandteil sind speziell konzipierte Praktika, zu deren Bearbeitung größtenteils Cloud-Dienste genutzt werden. Das Grundwissen für den Umgang mit den eingesetzten Cloud-Plattformen wird in einem initial durchgeführten Workshop zum Semesterstart vermittelt. Da es bisher an Sach- und Lehrbüchern auf dem Markt zur Ausarbeitung von Vorlesungen mit Cloud-Nutzung mangelt [ST14], wurde auf verschiedene andere Quellen zurückgegriffen. Das Basismodul wird in jedem Semester als Wahlfach angeboten, um allen Studierenden die Möglichkeit zu geben, dieses in Vorbereitung auf die Vertiefungsmodule zu besuchen. Während in der Vorlesung primär Konzepte und der mathematische Hintergrund der Verfahren thematisiert werden, untersuchen die Studierenden in einem kleinen Projekt im praktischen Teil des Moduls eine Fragestellung aus dem Forschungs- oder Industriealltag und beschäftigen sich mit der Anwendung der ML-Algorithmen. In den anschließenden Semestern kann das erworbene Wissen in verschiedenen anwendungsorientierten Themenschwerpunkten vertieft werden. Basierend auf der aktuellen Forschungstätigkeit im Studiengang wurden hier zunächst Vertiefungsmodule mit Schwerpunkt auf Robotics & Robot Vision, Data Science, Internet of Things und Natural Language Processing & Information Retrieval entwickelt. In diesen Vertiefungsmodulen steht die Untersuchung von praxisnahen Fragestellungen im Rahmen von Projekten im Vordergrund. Angestrebt sind also die Vermittlung methodischen Vorgehens zur effizienten Problemlösung und Optimierung angewandter Verfahren wie beispielsweise aktuell in einem Studierendenprojekt, Generative Adversarial Networks

zur Generierung von Trainingsdaten für die Klassifizierung von gefälschten Unterschriften. Den Studierenden ist es in den Vertiefungsmodulen freigestellt, in welcher Form sie die verschiedenen Cloud-Dienste in ihre Projekte einbinden. Darauf aufbauend kann eine Masterthesis in diesem Bereich erstellt werden, beispielsweise in Kooperation mit einem Firmenpartner oder im Kontext passender Forschungsprojekte am Campus Minden.

3 Cloud Konzept

Die Module aus dem Lehrkonzept haben potenziell sehr unterschiedliche Anforderungen hinsichtlich der nötigen Vorkenntnisse, der Auswahl der ML-Algorithmen und der gewünschten Abstraktionsebene. Wie schon beschrieben, liegt im Basismodul "Methoden des maschinellen Lernens" ein starker Fokus auf den Funktionsweisen und dem mathematischen Hintergrund der behandelten Methoden. Aus diesem Grund wird hier in der Regel auf den Einsatz etablierter Bibliotheken verzichtet und eine eigene Implementierung durch die Studierenden erwartet, um einen tiefen Einblick in die verschiedenen Algorithmen zu erlangen. Auf der anderen Seite stehen Module wie NLP, die stark anwendungsorientiert sind und deren Anforderungen eher auf der Anwendung, Evaluation und Kombination diverser Best Practices sowie (Hyper-) Parameter-Tuning von ML-Algorithmen liegen. Aus diesem Grund wurde die Entscheidung getroffen, mindestens zwei Cloud-Plattformen einzubeziehen, die sich im Service-Modell unterscheiden: Während das Machine Learning Studio (MLS) von Microsoft als Software-as-a-Service zu interpretieren ist, entspricht das Amazon Web Services-Angebot einem IaaS-Dienst. Die Integration von Cloud-Computing im beschriebenen Kontext lässt sich noch als Novum in der Lehre ansehen, daher existieren wenig bis keine Richtlinien für die effektive Nutzung und keine Orientierung bei der Erstellung eines Konzepts. Zwar wurde Ende 2018 eine Richtlinie für Cloud-Computing von einer internationalen Expertengruppe veröffentlicht [Fo18], jedoch muss diese noch sukzessive Einhalt in die Curricula finden und sich beweisen. Viele angesehene Hochschulen führen zwar das Schlagwort Cloud in der Prüfungsordnung, nutzen diese jedoch als Tool zu 80% nicht in der Praxis [ST14]. Ein Grund für die träge Zuwendung zu Cloud-Diensten in Unternehmen wie auch in Hochschulen ist das Mietmodell samt Nebeneffekten: Organisation der Datensicherung, -übernahme und -rückgabe sowie Lokalisierung zum Zeitpunkt der Verarbeitung, und die Stabilität der Mietkosten in absehbarer Zeit und Intervenierbarkeit seitens des Personals. Dabei gilt es insbesondere das Prinzip der Transparenz zu gewährleisten. Zur Anonymisierung sensibler Daten lassen sich beispielsweise Hashfunktionen einsetzen und bezüglich der Datensicherheit ein zentrales Schlüsselverwaltungssystem etablieren. Laut einer Studie von Bitkom Research [PV17] verbessern sich aber auch diverse Faktoren durch Einführung eines durchdachten Cloud-Computing-Konzeptes, darunter fallen die Aspekte Anwendungssicherheit, Implementierungszeiten und der IT-Administrationsaufwand. Entgegen der allgemeinen Meinung, dass Daten in der Cloud als unsicher gelten, zeigt die Studie deutlich, dass die Daten in der Public Cloud gegenwärtig als sicher beurteilt werden. Personenbezogene Daten sind jedoch zu vermeiden, da

dadurch wegen der Datenschutzgrundverordnung ein größerer bürokratischer Aufwand festzustellen ist. Die Informationen zum Empfänger, Zweck der Datenverarbeitung und die Aufbewahrung der Daten sind dokumentationspflichtig. Aufgrund der Tatsache, dass führende Cloud-Anbieter in den USA lokalisiert sind, ist besondere Vorsicht zu wahren. Unter diesen Voraussetzungen ist es schwer, direkte Angebote einzuholen. Durch das EU-US Privacy Shield-Datenschutzübereinkommen und die Partnerschaft mit Resellern, welche sich auf Öffentliche Einrichtungen und deren Ansprüche bei Support und Abwicklung spezialisiert haben, wird dieser Prozess jedoch deutlich vereinfacht. Die Verantwortung obliegt jedoch weiterhin der Hochschule, daher werden in Absprache mit den Resellern ausschließlich Prozesse implementiert und Dienste verwendet, welche die Bedingungen der DSGVO einhalten.

4 Evaluation

4.1 Methodik und Betrachtung der Ergebnisse

Zur Evaluation des Lehr- und Lernkonzepts wurde am Ende jedes Moduls eine Umfrage mit allen teilnehmenden Studierenden durchgeführt. Die Ergebnisse sind in Abb. 2 dargestellt. Den Umfrageergebnissen kann man entnehmen, dass die Studierenden mit dem inhaltlichen Aufbau der Module gut zurechtkamen.

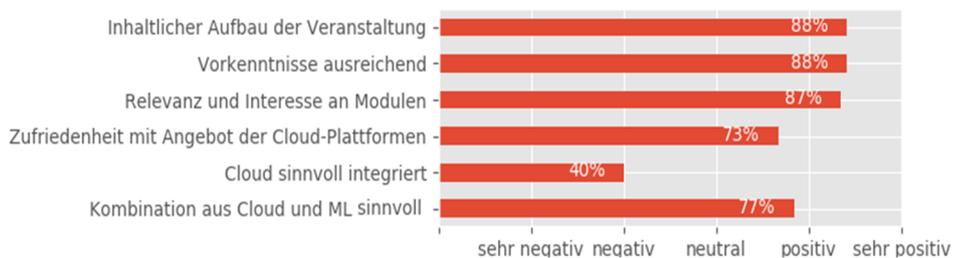


Abb. 2: Umfrageergebnisse

Es ist gelungen, das richtige Einstiegsniveau zu finden (88% der Teilnehmer gaben an, dass die Vorkenntnisse ausreichend waren zum Besuch der Module). Zudem zeigt sich, dass das Interesse an den Modulen wie erwartet sehr hoch ist und die vermittelten Inhalte für die Teilnehmer hohe Relevanz aufweisen. Den im Projekt verfolgten Ansatz, Rechenleistung flexibel über Cloud-Dienste anzumieten statt kosten- und wartungsintensive Hardware anzuschaffen, kann die überwiegende Mehrheit der Teilnehmer nachvollziehen (77%). Die Integration der Cloud-Dienste im Basismodul muss jedoch deutlich verbessert werden, lediglich 40% der Studierenden fanden den Einsatz der Cloud-Dienste im ersten Durchlauf gelungen. In Reaktion auf diese Umfrageergebnisse wurde bei der Wiederholung des Basismoduls eine hohe Priorität auf die stärkere und frühere Integration der Cloud-Dienste innerhalb des Semesters gelegt.

Beispielsweise fand eine Überarbeitung der Praktika statt, um mehr Aspekte der verschiedenen Dienste abzubilden. Außerdem findet zu Semesterstart nun ein Workshop als Pflichtveranstaltung statt, um den Studierenden den Zugang zu den verwendeten Cloud-Diensten zu erleichtern. Ob diese Änderungen sich positiv auswirken, wird sich bei den Umfragen nach dem aktuell laufenden Semester zeigen.

4.2 Evaluation der durch die Cloud-Plattformen entstandenen Kosten

Im Folgenden sollen die Kosten der eingesetzten Cloud-Dienste betrachtet werden, wobei zu beachten ist, dass pro Dienstleister eine maximale Anzahl von ca. 25 Studierenden das Angebot genutzt hat. Die Kosten der Cloud-Nutzung blieben dabei unter dem vorab kalkulierten Bedarf, da vor allem in der vorlesungsfreien Zeit wenig Nachfrage bestand. Für das Sommersemester 2018 sind so Kosten in Höhe von 2537,17 entstanden. Obwohl die hier betrachteten Dienste ähnliche Gesamtkosten erzeugt haben, setzen sich diese unterschiedlich zusammen. So fallen bei Microsoft Grundgebühren für jeden virtuellen Arbeitsplatz an. Dafür sind die Gebrauchskosten jedoch so gering, dass diese kaum ins Gewicht fallen. Bei AWS ist das Gegenteil der Fall (siehe Abb. 3).

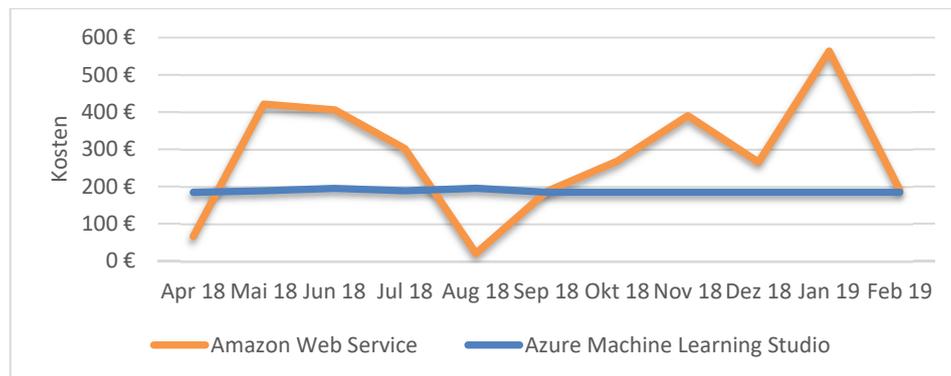


Abb. 3: Vergleich der Kosten pro Cloud-Dienst

Vergleicht man diese Kosten mit selbst angeschaffter Hardware, fallen diese sehr gering aus. Dies liegt hauptsächlich daran, dass im Lehreinsatz häufig stoßweise große Kapazitäten benötigt werden, beispielsweise während einer Lehrveranstaltung, es aber auch lange Zeiträume gibt, in der die Leistung nicht gebraucht wird. Die von uns überschlagenen Kosten, welche auf eingeholten Angeboten basieren, belaufen sich für einen Rechencluster mit vergleichbarer Leistung wie die von uns eingesetzten AWS Maschinen, bei 25 Teilnehmern auf ca. 3000 Euro im Monat. Dabei sind relevante Kostenfaktoren wie Strom und Wartung der Cluster noch nicht eingeschlossen, die jedoch auch einen erheblichen Anteil der Gesamtkosten ausmachen [YF15].

5 Zusammenfassung und Ausblick

Das in diesem Bericht beschriebene Konzept zur Ausbildung im Bereich Maschinelles Lernen wurde erfolgreich umgesetzt. Das Basismodul und verschiedene Vertiefungsmodule wurden erstmalig durchgeführt und sind auf große Resonanz unter den Studierenden des Masterstudiengangs Informatik am Campus Minden gestoßen. Im Projekt konnten wertvolle Erfahrungen im Einsatz von Cloud-Diensten für Maschinelles Lernen in der Lehre gesammelt werden, die in die Verbesserung der Module einfließen werden. Die daraus resultierenden Vorteile gegenüber der Anschaffung eigener Hardware, wie Wartungsfreiheit und verfügbare Ressourcen, stachen insgesamt hervor. Durch die Einführung von Cloud-Diensten kann die Qualität und Kosteneffektivität von Maschinellem Lernen im Lehreinsatz zunehmen, sowie der Energieverbrauch durch die dezentrale Virtualisierung abnehmen. Bisher wurden die Plattformen AWS und Azure eingesetzt. Im aktuell laufenden Semester wird das Angebot um die GCP ergänzt. Ein wichtiger Aspekt ist dabei die Kostenkontrolle und -vorhersage. Im bisherigen Verlauf dieses Projekts sind die Kosten unter den Prognosen geblieben. Speziell bei der Nutzung der GCP erscheint es sinnvoll zu evaluieren, inwieweit isolierte Lernumgebungen zum kostenlimitierten Experimentieren durch Cloud-Sandboxes eingesetzt werden können. Alternativ würde ein eigenes Management der Ressourcen in Verbindung mit einer restriktiven Konfiguration des Cloud Identity & Access Management eine sichere und isolierte Umgebung für die Cloud-Dienste darstellen.

Danksagung

Die vorliegende Arbeit wurde finanziell im Rahmen des BMBF-Vorhabens "ML Camp - Machine Learning Campus Minden" unter dem Kennzeichen 01 IS 17081 gefördert.

Literaturverzeichnis

- [Fo18] Foster, Derek, et al.: Cloud computing: developing contemporary computer science curriculum for a cloud-first future. Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. ACM, 2018.
- [YF15] Yaghmaei, Omid; Binesh, Fatemeh: Impact of applying cloud computing on universities expenses. IOSR Journal of Business and Management Volume 17.2 S. 42-47, 2015.
- [PV17] Pols, Axel; Vogel, Marko: Bitkom Research GmbH, <https://www.bitkom.org/sites/default/files/pdf/Presse/Anhaenge-an-PIs/2017/03-Maerz/Bitkom-KPMG-Charts-PK-Cloud-Monitor-14032017.pdf>, Stand: 05.03.2019.
- [ST14] Scherzinger, Stefanie; Thor, Andreas: Cloud-Technologien in der Hochschullehre – Pflicht oder Kür?. Datenbank-Spektrum 10/14 S. 131-134, 2014.