**K O N Z E P T**

**für**

**Python Bootcamp für Medizinstudenten und verwandte Studiengänge**

Änderungsverzeichnis

| Datum | Version | Änderung | Autor |
| --- | --- | --- | --- |
| 06.07.2023 | 0.1 | Beginn der Erstellung eines Konzepts | E. Einspänner |
| 21.07.2023 | 0.2 | Konkretisierung aller Inhalte | E. Einspänner |
| 11.08.2023 | 1.0 | Ergänzungen nach Besprechung | E. Einspänner |
|  |  |  |  |
|  |  |  |  |

Inhalt

[**1.** **Einleitung / Rahmenbedingung** 3](#_Toc142651720)

[**1.1** **Auftrag** 3](#_Toc142651721)

[**1.2** **Planungshorizont** 3](#_Toc142651722)

[**2.** **Ausgangslage** 4](#_Toc142651723)

[**3.** **Zielsystem** 5](#_Toc142651724)

[**3.1** **Ziele und Lerninhalte** 5](#_Toc142651725)

[**3.2** **Zielgruppen** 6](#_Toc142651726)

[**3.3** **Erfolgskontrolle** 6](#_Toc142651727)

[**4.** **Strategie und Massnahmen** 7](#_Toc142651728)

[**4.1** **Plattform** 7](#_Toc142651729)

[**4.2** **Zielgruppen** 7](#_Toc142651730)

[**5.** **Projektmanagement** 8](#_Toc142651731)

[**5.1** **Organisationsaufbau** 8](#_Toc142651732)

[**5.2** **Organisationsablauf** 8](#_Toc142651733)

[**6.** **Zukunftsvision** 9](#_Toc142651734)

# **Einleitung / Rahmenbedingung**

# **Auftrag**

Erstellung eines Python Bootcamps für Medizinstudenten als Wahlveranstaltung, sowie weiterer Studiengänge mit medizinischem Hintergrund (z.B. Medizintechnik, -informatik).

# **Planungshorizont**

Die Fertigstellung einer ersten Version (v1.0) des Bootcamps ist bis Anfang 2024 geplant. Dies beinhaltet die Bereitstellung der grundlegenden (interaktiven) Python Notebooks sowie der dazugehörigen Slides, welche die Notebooks didaktisch erweitern.

Das SS 2024 dient als erster Test des Bootcamps, sowohl inhaltlich als auch funktionell. Nach anschließender Evaluierung, Auswertung und ggf. Anpassungen ist eine Ausweitung des Bootcamps möglich. Dies würde auch Student:innen außerhalb der OvGU den Zugang ermöglichen.

# **Ausgangslage**

Derzeit existieren keine Lernangebote bzw. –inhalte die einen praxisnahen Einstieg in die Verwendung der Programmiersprache Python innerhalb des Klinikalltags bereitstellen. Die wachsenden Anforderungen an die Studierenden hinsichtlich der Datenbeschaffung, -aufarbeitung, -analyse und der Gewinn von Wissen aus den Daten ist eine zunehmend wichtige Kernkompetenz.

Im Rahmen von Abschluss- und Promotionsarbeiten, und darüber hinaus bei klinischen Forschungsprojekten, ist die Arbeit mit vielen Datensätzen unerlässlich. Für eine Aus- und Aufarbeitung der Daten für die eigentliche Forschungsfrage nutzen die Student:innen oftmals Tools wie Microsoft Excel, um Ihre Daten manuell zu verwalten. Diese Arbeitsweise ist zum einen sehr zeitintensiv und zum anderen sehr fehleranfällig. Weiterhin müssen solche manuell angelegten «Datenbanken» regelmäßig manuell up-to-date gehalten werden.

Mit geeigneten Pipelines, programmiert in Python, lassen sich Datensätze schnell erfassen und sind in geeigneter Form direkt weiterverwendbar für etwaige Datenanalysen. Solche Pipelines sind semi- bis vollautomatisch integrierbar und sparen den Student:innen wertvolle Zeit. Effizienz und Fehlerwahrscheinlichkeiten nehmen ab. Jedoch erfordert die Programmierung von komplexeren Tools in Python gewisse Vorkenntnisse. Hierfür besteht derzeit kein adäquates Angebot seitens der Universität bzw. der Medizinischen Fakultät.

Viele digitale Angebote von Unternehmen sind mit hohen Kosten verbunden und decken meist nicht die relevanten Lerninhalte, besonders mit Hinblick auf medizinische Fragestellungen, ab. Aus diesem Grund soll ein kostenloser, open-source Kurs (Bootcamp) entworfen werden, welcher die Lernenden einen Einstieg in Python ermöglicht.

Alle Lerninhalte werden in der englischen Sprache veröffentlicht. Der Hintergrund ist, möglichst alles Student:innen das Bootcamp anbieten zu können. Zudem ist Englisch die Wissenschaftssprache und ebenfalls eng mit der Programmierung verknüpft.

# **Zielsystem**

# **Ziele und Lerninhalte**

Durch die zunehmende Digitalisierung der Medizin und des raschen Voranschreitens der Entwicklung und Integration von Machine Learning ist eine Weiterbildung von Student:innen in die Grundlagen der Datenanalyse und –verarbeitung, sowie die Extraktion von Wissen aus Daten (Data Science) notwendig.

Dabei soll der Fokus weniger auf der Vermittlung mathematischer Grundlagen liegen, sondern vorrangig auf der Anwendung verschiedenster Module mit Hinblick auf die Problemstellung, welche sich aus der Klinikroutine und Forschung ergeben.

Als Programmiersprache wird hierfür Python verwendet. Python ist eine benutzerfreundliche und einfach zu lernende Programmiersprache, mit einer großen Bandbreite an Anwendungen und einer großen Community. Zudem ist Python Open-Source, plattformübergreifend und kostenfrei. Weiterhin ist Python eine der am häufigsten verwendeten Programmiersprachen für Data Science und Machine Learning.

Neben der Bereitstellung interaktiver Notebooks sollen unterstützende Lerninhalte (in Form von Slides) bereitgestellt werden.

Folgende Lerninhalte, jeder Schwerpunkt stellt dabei einen separaten Skill Track dar, werden verfolgt:

* Einführung in die Programmiersprache Python
* Einführung in die grundlegenden Konzepte der Programmierung
* Vorstellung wichtiger Python-Bibliotheken und deren Verwendung im Bereich Data Science
  + Numpy
  + Matplotlib
  + Seaborn
  + SciPy
  + Statsmodels
  + Scikit-learn
* Vorstellung wichtiger Python-Bibliotheken und deren Verwendung im Bereich DICOM und Bildverarbeitung (ergänzend zu den o.g.)
  + Pydicom
  + OpenCV
* Einführung in Datenbanken
  + Erstellung
  + Abfragen (Query’s), Verwaltung
  + Datenanalyse und –visualisierung mittels Pandas
  + XNAT
* Grundlagen der Statistik
  + Erstellung von Statistiktools
  + Anwendung auf Teil- und Gesamtdatensatz
* Einführung in Machine Learning
  + Tensorflow/Keras
  + PyTorch

Weiterhin sollen «geführte Notebooks» für kleinere Projekte erstellt werden, in denen die Student:innen das Gelernte der verschiedenen Skill Tracks anwenden können. Nachfolgend soll kurz anhand eines Beispiels der Ablauf eines solchen Projektes dargestellt werden:

* Datensatz: mehrere Patientenaufnahmen (CT oder. MRT) bspw. vom Hirn (mit Pseudo-DICOM-Headern)
* Verschiedene Aufgaben:
  + Einlesen und überprüfen der .dcm-Files
  + Fehlerhafte .dcm-Header identifizieren
  + Alle dcm.-Files anonymisieren
  + Sub-Plots (beispielhaft) der Bilddaten erstellen
  + alle Bilddatensätze normalisieren
  + automatische Segmentierung einer spezifischen Region
  + Auswertung der segmentierten Hirnregion
  + Visualisierung der Ergebnisse

# **Zielgruppen**

Die Zielgruppe des Python Bootcamps sind Student:innen der Medizinischen Fakultät der OvGU. Dabei sollen die Inhalte stets praxisnah entworfen werden. Durch Weiterentwicklung und die Integration mehrerer Anforderungsprofile wäre eine Ausweitung auf weitere Studiengänge mit medizinischem Bezug (z.B. Medizintechnik) möglich.

# **Erfolgskontrolle**

Um den Teilnehmer:innen ein Abschluss bzw. Teilnehmerzertifikat aushändigen zu können ist ein Abschlussprojekt vorgesehen. In diesem sollen die Student:innen eine Aufgabe mit ihrem erworbenen Wissen selbstständig lösen und ihr Ergebnis einschicken. Nach Sichtung der Ergebnisse wird ein entsprechendes Zertifikat für die Teilnahme und den erfolgreichen Abschluss ausgehändigt.

# **Zeitaufwand**

Das Bootcamp soll als reines Onlineangebot angeboten werden. Es werden keine festen Timeslots für einzelne Lerninhalte vorgeschrieben, die Student:innen können somit in einem eigens festgelegtem Tempo die einzelnen Skill Tracks absolvieren.

Der zeitliche Aufwand wird für einen Anfänger mit moderatem Lerntempo aus 20 bis 30 Stunden pro Semester geschätzt, dies entspricht ca. 2 SWS pro Semester.

# **Strategie und Massnahmen**

# **Plattform**

Für die Umsetzung der Ziele und einen einfachen Einstieg ohne Hürden, entstehend durch Installations- und Einrichtungsvorgänge, wird eine webbasierte Bereitstellung der Inhalte bevorzugt.

Um es den Studenten möglichst leicht zu gestalten sollen «elektronische Notizbücher», s.g. Notebooks, genutzt werden. Dies ermöglicht einen interaktiven Zugang vom Code zum Programm, von Daten zur Visualisierung und vom Text zur Erläuterung. Hierfür bieten sich zwei möglich Alternativen an: Jupyter Notebooks oder Google Colab.

Jupyter Notebooks ist eine non-profit Open-Source-Möglichkeit, welche lokale Maschinenressourcen für die Codeausführung nutzt. Für einen reibungslosen Ablauf wären eine Installationsbeschreibung und die Erstellung einer Konfigurationsdatei notwendig.

Google Colab bietet eine vollwertige Cloud-App für die Python Codierung und kann direkt in den Webbrowsern (z.B. Google Chrome oder Mozilla Firefox) genutzt werden. Hierfür müssen keine lokalen Installationen durchgeführt werden.

Die Tabelle 1 verdeutlicht die Vor- und Nachteile der beiden Plattformen.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Jupyter Notebook** | **Google Colab** |
| Hintergrund | Non-profit,  Open-Source | Google |
| Cloud-basiert | ❌ | ✔ |
| Dateisynchronisierung | ❌ | ✔ |
| Gemeinsame Nutzung von Dateien | ❌ | ✔ |
| Installation | ✔ | ❌ |
| Dateiansicht ohne Installation | ❌ | ✔ |

Tabelle 1 Vor- und Nachteile von Jupyter Notebook und Google Colab

# **Support**

Ein zentraler Bestandteil ist die Bereitstellung eines geeigneten Supportsystems. Hierfür werden verschiedene Wege/Methoden zur Verfügung gestellt:

* FAQ/Wiki auf GitHub/GitLab
* README.md pro Skill Track und allgemeine Anleitungen
* Issues bei techn. Fragen/Anmerkungen/Verbesserungvorschlägen
* Mailkontakt (zentrale Mailadresse) bei inhaltlichen Frage, Anmerkungen usw.
* eine regelmäßig stattfindende Livesession (1h, per Zoom o.ä.)

# **Evaluation**

Jeder Student:innen muss vor und nach dem Bootcamp ein Fragebogen zur Evaluation des Bootcamps, sowie zur Bestimmung des Lernerfolgs ausfüllen. Mithilfe dieses Fragebogens sollen folgende Fragen beantwortet werden:

* Welcher Lernfortschritt wurde erzielt?
* Wie werden Aufgaben zukünftig mittels Python umgesetzt?
* Gesteigerte Effizienz?
* Hatte das Bootcamp eines Einfluss auf die Arbeitsweise bzw. Herangehensweise der Student:innen?

Weiterhin soll die Auswertung für die weitere Verbesserung des Bootcamps genutzt werden.

# **Projektmanagement**

# **Organisationsaufbau**

Das Projekt wird von der Klinik der Neuroradiologie (vertreten durch Prof. D. Behme, Dr. S. Klebingat und M.Sc. E. Einspänner) und Prof. S. Saalfeld (TU Ilmenau, Medizinische Informatik) betreut.

Die Verwaltung und gemeinsame Kollaboration läuft über GitHub/GitLab, eine Webanwendung zur Versionsverwaltung für Softwareprojekte auf Git-Basis. Hierüber ist ein gemeinsames arbeiten und weiterentwickeln der Inhalte möglich.

# **Organisationsablauf**

Die Bereitstellung neuer Lerninhalte sowie die Aktualisierung bestehender Inhalte erfolgt über das zentrale Repository auf GitHub/GitLab.

Der Inhalt einzelner Notebooks sowie alle weiteren Inhalte durchlaufen vor Veröffentlichung einer Qualitätskontrolle. Eine jede Qualitätskontrolle beinhaltet folgende Kriterien:

* Anforderung / Skill Level
* Praxisbezug
* Datenqualität
* Prüfung auf möglichen Lerneffekt

# **Zukunftsvision**

Um fortlaufend das Lernangebot zu erweitern sollen zukünftig folgende Features hinzugefügt werden:

* Challenges – 1-wöchige Gruppenarbeit zur Lösung eines realen Problems