

1 Szenario A: Neuinstallation auf einem eigenen Server

Beim Aufsetzen einer vollständigen Umgebung und dem Erzeugen eines neuen Modells sind die folgenden Schritte erforderlich.

1.1 Installation der erforderlichen Software und Bibliotheken

- Die Basisinstallation der erforderlichen Bibliotheken (wie *numpy*, *pandas*, *scikit-learn*, *nltk*, etc.) wird im Dokument *AI_Model_Dokumentation.md* beschrieben.
- Es wird empfohlen, die im Dokument genannten Bibliotheken und Python-Version auf dem Server zu installieren.

1.2 Datenverarbeitung und Vorbereitung

- Die vorbereitenden Schritte sind in *Datenverarbeitung_Dokumentation.md* beschrieben. Diese Schritte sollten vor der Modellerstellung durchgeführt werden, um die Rohdaten bereitzustellen und in die korrekte Struktur zu bringen.
- Wesentliche Schritte umfassen die Nutzung der CSV-Dateien und die Bereinigung, etwa mit dem *NullToZero*-Generator (siehe Punkt 1.3).

1.3 Bereinigung der Daten

- Der Prozess des *NullToZero*-Generators (*NullToZero_Generator.ipynb*, siehe auch *NullToZero_Generator_Dokumentation.md*) sorgt dafür, dass alle leeren Zellen in der Matrix korrekt mit Nullen gefüllt werden, sodass ein einheitliches Datenformat für das Modell entsteht.
- Dieser Schritt ist vor der Modellerstellung erforderlich, da er die Qualität der Daten sicherstellt.

1.4 Erstellen des AI-Modells

- Das eigentliche Training und die Erstellung des Modells erfolgt mit dem in *AI_Model_Dokumentation.md* beschriebenen Skript *tfidf_python_Category.ipynb* (inkl. Kategoriefilter).
- Wenn das Modell erfolgreich trainiert ist, wird es als Datei *tfidf_model.pkl* gespeichert.

1.5 Modellbereitstellung und Speicherung

- Speichern und Bereitstellen des Modells erfolgen gemäß den Anweisungen in *AI_Model_Dokumentation.md*. Das trainierte Modell *tfidf_model.pkl* ist die Grundlage für Empfehlungen und Analysen.
- Diese Schritte stellen sicher, dass das Modell in produktiven Umgebungen abrufbar und nutzbar ist.

2 Szenario B: Update des KI-Modells bei geänderten Daten

Wenn die KI-Umgebung bereits installiert ist, sich aber die Metadaten geändert haben, können die Schritte verkürzt werden, um lediglich das Modell zu aktualisieren.

2.1 Daten bereinigen und aktualisieren

- Führen Sie das *NullToZero_Generator.ipynb*-Notebook aus, um die neuen Metadaten im vorhandenen Format zu bereinigen.
- Diese Schritte sind kurz in *NullToZero_Generator_Dokumentation.md* erläutert.

2.2 Modell neu trainieren

- Das erneute Training des Modells wird anhand der aktualisierten und bereinigten Daten in *AI_Model_Dokumentation.md* beschrieben.
- Es ist nur der Modelltrainingsabschnitt erforderlich, Schritte zur Installation und zum Setup entfallen.

2.3 Modell speichern und überschreiben

- Nach dem Training wird das bestehende Modell *tfidf_model.pkl* durch das aktualisierte Modell überschrieben, ohne dass die Infrastruktur angepasst werden muss.