OTH AMBERG-WEIDEN
Fakultät Elektrotechnik, Medien und Informatik
Prof. Dr. F. Brunner
Wintersemester 2021/2022

Ostbayerische Technische Hochschule
Amberg-Weiden

Projektarbeit „Machine Learning“

Ausgabe: 10.01.2023 Deadline: 05.02.2023

Ziel der Projektarbeit

In der vorliegenden Projektarbeit sollen Machine Learning-Ansätze für die vorausschauende Wartung („predictive maintenance“) von U-Bahn-Fahrzeugen entwickelt und evaluiert werden.

Modalitäten

- Die Bearbeitung der Projektarbeit erfolgt in der Programmiersprache Python unter Verwendung der Machine Learning-Bibliothek Scikit-learn.
- Als Ergebnis ist ein Jupyter-Notebook namens `nachname_vorname.ipynb` zu erstellen und elektronisch via Moodle abzugeben, das den vollständigen Programmcode sowie die gesamte Dokumentation und die Ergebnisse (insb. eingebettete Grafiken und erläuternden Text) enthält. Achten Sie auf eine ansprechende und übersichtliche Darstellung der Inhalte, indem Sie beispielsweise die Strukturierungsmöglichkeiten der Markdown-Sprache nutzen.
- Der eingereichte Programmcode soll auf den Rechnern des GPU-Labors mit den dort installierten Versionen von Anaconda/Python/Scikit-learn lauffähig sein. Geben Sie im Kopf des Dokuments an, ob und ggf. welche zusätzlichen Pakete installiert werden müssen.
- Die Bearbeitung der Projektarbeit ist in Gruppen von maximal zwei Personen zulässig. Im Fall einer Zweierabgabe genügt es, wenn ein Gruppenmitglied das Jupyter-Notebook elektronisch einreicht. Im Kopf des Dokuments sind alle Gruppenmitglieder zu nennen.
- Die Abgabe des Notebooks und der Präsentationsfolien hat bis spätestens 05.02.2023 um 23:59:59 Uhr über Moodle zu erfolgen.
- Die Vorstellung der Ergebnisse in Form einer ca. 20-minütigen Präsentation erfolgt am 07.02.2023. Zur Planung und Einteilung der Vorträge tragen Sie sich bitte bis spätestens 16.01.2023 in eine Umfrage in Moodle ein.
- Die auf der letzten Seite abgedruckte Erklärung ist von jedem Gruppenmitglied auszufüllen und in gescannter Form mit den übrigen Dokumenten in Moodle hochzuladen. Im Fall einer Zweiergruppe kann die Erklärung auch von jedem Gruppenmitglied einzeln ausgefüllt und hochgeladen werden.

Anforderungen und Bewertungsgrundlagen

Code und Ergebnisdokumentation

- Der Code ist klar strukturiert, gut lesbar, nachvollziehbar und ausreichend kommentiert.
- Der Code ist effizient und greift, wo möglich, auf vorhandene Funktionalität von Python, NumPy, Pandas und Scikit-learn zurück.
- Das eingereichte Jupyter-Notebook ist lauffähig und übersichtlich gestaltet. Nutzen Sie dazu die Strukturierungsmöglichkeiten der Markdown-Sprache innerhalb von Jupyter-Notebooks. Versehen Sie das Notebook mit einer Gliederung, die die einzelnen Aufgaben bzw. Arbeitsschritte verlinkt, einer Zusammenfassung und einem Quellenverzeichnis.
- Neben dem Programmcode sollen im Notebook die einzelnen Arbeitsschritte und Ergebnisse ausführlich dokumentiert werden. Beschreiben Sie dabei nicht nur, „was“ Sie gemacht haben, sondern vor allem „warum“.

Qualität und Leistungsfähigkeit der Modelle

- Die entwickelten Modelle besitzen eine hinreichende Güte.
- Diese soll anhand geeigneter Kriterien und Techniken gemessen, optimiert und bewertet werden.

Methodik und Vorgehen

- Das Vorgehen zur Modellerstellung und -bewertung ist fachlich begründet und nachvollziehbar.
- Die einzelnen Arbeitsschritte sind methodisch korrekt ausgeführt worden.

Ergebnispräsentation

- Die Ergebnisse der Projektarbeit sind in einem 20-minütigen Vortrag zu präsentieren.
- Zu diesem Zweck soll ein Foliensatz mit maximal 15 Folien erstellt werden. Dieser geht als Teil der Dokumentation in die Bewertung ein.

Aufgabenstellungen

In der vorliegenden Projektarbeit sollen im Rahmen der folgenden Aufgaben verschiedene Prognosemodelle erstellt und evaluiert werden, die für die vorausschauende Wartung (engl. „predictive maintenance“) von U-Bahn-Zügen eingesetzt werden können.
Durch den Einsatz geeigneter Modelle soll dabei eine rechtzeitige Wartung von Zügen planmäßig durchgeführt werden, um ungeplante Ausfälle zu vermeiden, die sich nicht zuletzt auf die Fahrgäste negativ auswirken.
Die Modellbildung soll auf der Basis des sog. „`MetroPT`“-Datensatzes erfolgen. Dieser Datensatz aus dem Jahr 2022 stammt vom U-Bahn-Netz in der portugiesischen Stadt Porto und wurde als Benchmark-Datensatz für Predictive Maintenance veröffentlicht, vgl. [1]. Er enthält Sensordaten der sog. „Air Producing Unit“ (kurz: APU) eines Zugs für den Zeitraum Januar bis Juni 2022. Bei der APU handelt es sich um eine Systemkomponente des Zugs, die im laufenden Betrieb verschiedene wichtige Funktionen erfüllt, und deren Ausfall eine sofortige Außerbetriebnahme und Reparatur erforderlich macht. Weiterhin werden Angaben zu drei Störungsfällen gemacht, die sich während des o.g. Betrachtungszeitraums ergeignet haben. Diese können verwendet werden, um geeignete Zielvariablen für Methoden des Supervised Learning abzuleiten.
Eine ausführliche Beschreibung des Datensatzes entnehmen Sie bitte der Veröffentlichung [1]. Der Datensatz ist unter folgendem Link zum Download verfügbar:

`https://zenodo.org/record/6854240#.Y7NB9RWZOHs`

Aufgabe 1 (Klassifikation des Systemzustands)
Erstellen Sie binäre Klassifikationsmodelle zur Vorhersage des aktuellen Zustands der APU auf der Basis der gemessenen Sensordaten. In dieser Aufgabe soll zunächst nur danach differenziert werden, ob die APU in Ordnung oder nicht in Ordnung ist (binäre Klassifikation). Verwenden Sie dazu aus den in der Vorlesung besprochenen Verfahren mindestens

- einen nicht-parametrisierten Modellansatz,
- einen parametrisierten Modellansatz,
- ein Verfahren aus dem Bereich des Ensemble Learning.

Erstellen Sie zunächst einen Trainingsdatensatz, z.B. durch geeignete Datentransformationen, Feature Engineering und ggf. Feature Extraction. Berücksichtigen Sie bei der Modellierung die sequentielle Struktur der Daten. Wenden Sie zur Modellerstellung (in dieser und den folgenden Aufgaben) geeignete Maßnahmen und Techniken an, damit die resultierenden Modelle eine möglichst hohe Güte aufweisen und beurteilen Sie diese anhand geeigneter Kriterien. Modularisieren und automatisieren Sie Ihren Workflow, damit die einzelnen Schritte in den folgenden Aufgaben ggf. wiederverwendet werden können. Achten Sie darauf, dass Ihre Modelle auf unbekannte Daten angewendet werden können, die ggf. fehlende Werte enthalten, auch wenn der gegebene Datensatz keine fehlenden Werte enthält. Welche Features erweisen sich als besonders aussagekräftig für die gegebene Aufgabenstellung?

Aufgabe 2 (Vorhersage des Eintretens von Störungen)
Erstellen Sie nun Klassifikationsmodelle, um anhand der gegebenen Sensormessdaten vorherzusagen, ob innerhalb eines bestimmten Zeitraums (z.B. 1 Stunde, 2 Stunden etc.) eine Störung der APU auftreten wird. Laut Betreiber wäre es wünschenswert, mindestens zwei Stunden im Voraus eine Störung vorhersagen zu können, um rechtzeitig Maßnahmen einzuleiten, vgl. [1]. Testen Sie verschiedene Prognosezeiträume und stellen Sie die resultierenden Modelle gegenüber.

Aufgabe 3 (Vorhersage der Dauer von Störungen)
Entwickeln Sie Prognosemodelle zur Vorhersage der Störungsdauer und beurteilen Sie auf geeignete Weise deren Güte sowie deren Eignung für den Einsatz in der Praxis. Sofern diese aus Ihrer Sicht nicht ausreichend ist, skizzieren Sie Maßnahmen, durch die die Güte verbessert werden könnte.

Aufgabe 4 (Vorhersage der gestörten Komponente)
Untersuchen Sie, ob sich der Datensatz auch dazu eignet, die von einer Störung betroffenen Komponente anhand der Sensordaten zu identifizieren. Erstellen und evaluieren Sie dazu entsprechende Modelle.

Aufgabe 5 (Störungserkennung mit Hilfe von Unsupervised Learning)
Eine Herausforderung bei der Modellbildung für Predictive Maintenance ist häufig das Fehlen von Informationen zu historischen Störungen, sodass Ansätze des Supervised Learning nicht anwendbar sind. In diesem Fall können Methoden des Unsupervised Learning eine Option sein.
Wenden Sie auf den Datensatz aus Aufgabe 1 (ohne Labels) ein Clustering-Verfahren an und überprüfen Sie anhand der gegebenen Informationen zu den historischen Systemausfällen, ob und wie gut sich durch einen solchen Ansatz Ausnahmezustände (Anomalien bzw. Störungen) von „normalen“ Systemzuständen der APU unterscheiden lassen.

Literatur

[1] B. Veloso, R. P. Ribeiro, J. Gama, and P. M. Pereira, “The MetroPT dataset for predictive maintenance,” *Scientific Data*, vol. 9, no. 1, pp. 1–8, 2022.

Anlage zur Projektarbeit „Machine Learning“

Wintersemester 2022/2023

Prof. Dr. Fabian Brunner

Name, Vorname Gruppenmitglied 1:

Matrikelnummer Gruppenmitglied 1:

Name, Vorname Gruppenmitglied 2:

Matrikelnummer Gruppenmitglied 2:

Erklärung

Hiermit wird erklärt, dass die eingereichte Projektarbeit ausschließlich von den o.g. Personen erstellt wurde. Alle verwendeten Hilfsmittel und Quellen sind in der Arbeit angegeben worden.

Ort, Datum

Unterschrift