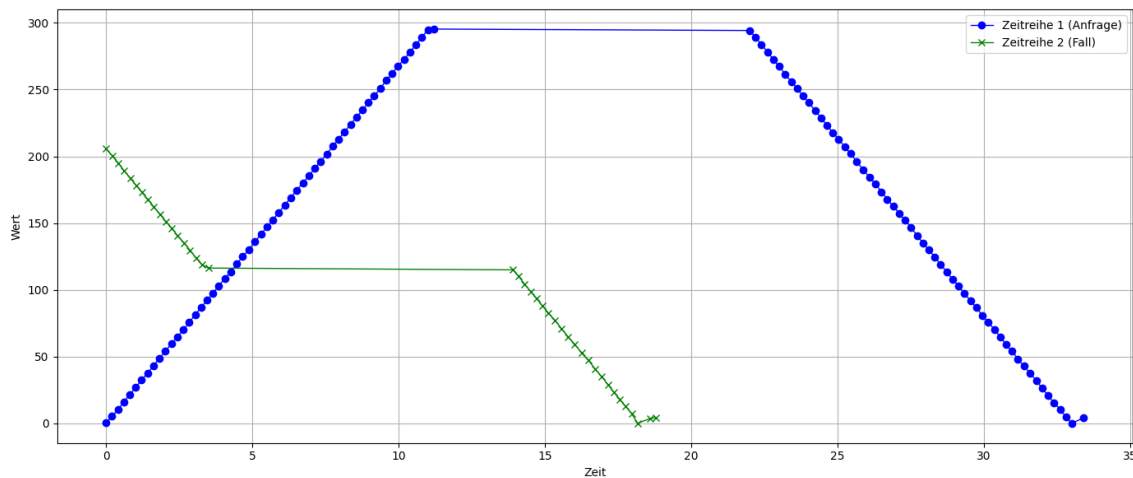


Evaluation von Erklärtexten

Die Grafik zeigt zwei Zeitreihen aus ProCAKE – W19 (Anfrage) und W25 (Fall). Der nachfolgende Text wurde von ChatGPT generiert und beschreibt die Ähnlichkeit der Zeitreihen mithilfe des Dynamic Time Warping (DTW)-Verfahrens. Ziel dieser Evaluation ist es, die Qualität der generierten Erklärung einzuschätzen. Bewertet werden dabei die fachliche Korrektheit, die inhaltliche Vollständigkeit sowie die Relevanz der Informationen in Bezug auf den konkreten Fall. Darüber hinaus werden die Verständlichkeit für fachfremde Nutzer:innen, logischer Aufbau und der sprachliche Lesefluss berücksichtigt.

Bitte lesen Sie die Erklärung sorgfältig durch und beantworten Sie anschließend die Bewertungsfragen anhand der bereitgestellten Skala. Ergänzend können Sie offene Anmerkungen zu möglichen Fehlern, Verständlichkeitsproblemen oder Verbesserungsvorschlägen machen.



LLM Output:

In diesem Beispiel werden zwei Zeitreihen gegenübergestellt, eine Anfrage-Zeitreihe mit 108 Messpunkten über etwa 33,4 Sekunden und eine Fall-Zeitreihe mit 41 Messpunkten über rund 18,8 Sekunden. Die Messpunkte sind sowohl in ihrer Anzahl als auch in ihren Zeitabständen unterschiedlich verteilt, was bereits im Diagramm daran sichtbar ist, dass die blaue (Anfrage) Reihe wesentlich länger und dichter besetzt ist als die grüne (Fall) Reihe.

Zum Vergleich der beiden Reihen wurde der Dynamic Time Warping Algorithmus (DTW) verwendet. Dieser Algorithmus ordnet die Messpunkte beider Reihen so zu, dass zeitliche Verschiebungen oder Unterschiede in der Geschwindigkeit ausgeglichen werden. Dafür wird zunächst eine Matrix erstellt, in der jeder Punkt der ersten Zeitreihe mit jedem Punkt der zweiten verglichen wird. An jeder Stelle bewertet der Algorithmus, wie ähnlich sich zwei Messpunkte sind; dabei spielen sowohl die Differenz der Messwerte als auch der Abstand ihrer Zeitstempel eine Rolle, wobei der Messwert deutlich stärker gewichtet wird.

Die lokale Ähnlichkeit zwischen zwei Punkten nimmt linear ab, je größer der Unterschied im Wert (bis zu 20 Einheiten) oder der Zeit (bis zu 100 Sekunden) ist. Treffen zwei Messungen genau im gleichen Wert und zum selben Zeitpunkt aufeinander, ist die Ähnlichkeit maximal, bei starkem Unterschied fällt sie auf null zurück.

Um den bestmöglichen Vergleichspfad zu finden, sucht DTW in der Matrix immer den Weg, der in Summe die höchsten Ähnlichkeiten aufweist – bevorzugt über direkte Zuordnungen (diagonale Schritte), die doppelt gewichtet werden. Alternativ sind auch horizontale oder vertikale Schritte möglich, die zeitliche Verschiebungen oder Lücken ausgleichen.

Im vorliegenden Fall zeigt sich im DTW-Pfad eine starke Verzerrung am Anfang: Über 70 Messpunkte aus der Anfrage-Zeitreihe werden ausschließlich dem ersten Wert der Fall-Zeitreihe zugeordnet. Dieser Bereich wird im DTW durch horizontale Schritte dargestellt und entsteht, weil in diesem Abschnitt keine passenden Wertepaare existieren. Erst ab einem Wert von etwa 200 in der Anfrage-Zeitreihe finden sich vergleichbare Werte in der Fall-Zeitreihe. Ab diesem Punkt beginnt der DTW-Pfad, einzelne Messpunkte zunehmend direkt zuzuordnen, sodass die Zuordnungspunkte nun überwiegend diagonal verlaufen. In diesem Bereich sind die Werte beider Reihen ähnlich, dennoch besteht ein deutlicher Zeitversatz.

Der berechnete DTW-Ähnlichkeitswert beträgt 0,55. Dieser Wert liegt im mittleren Bereich der Skala und spiegelt wider, dass es einerseits Bereiche mit ähnlichen Werten gibt, die jedoch zeitlich deutlich verschoben sind. Insgesamt überwiegen die Unterschiede im Verlauf, insbesondere zu Beginn und in der Plateau-Phase, in denen kaum Übereinstimmungen vorliegen und fast keine direkten Punktpaare gebildet werden können. Die gemessene Ähnlichkeit ist daher vor allem auf den abfallenden Abschnitt zurückzuführen, in dem die Zeitreihen trotz Zeitverschiebung ähnliche Werte erreichen.

Bitte bewerten Sie die folgenden Aussagen auf einer Skala von 1 bis 5:

1 = stimme überhaupt nicht zu 5 = stimme voll zu

1. Die Erklärung ist fachlich korrekt - alle dargestellten Fakten, Begriffe, Interpretationen und DTW-spezifischen Konzepte sind richtig und frei von sachlichen Fehlern.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

2. Sowohl die relevanten Zeitreihenmerkmale als auch die Funktionsweise des DTW-Algorithmus (Matrix, Pfad, lokale und globale Ähnlichkeit) werden vollständig beschrieben.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

3. Die Erklärung beschränkt sich auf zentrale Aspekte, die für das Verständnis des konkreten DTW-Vergleichs wichtig sind, und vermeidet allgemeine oder irrelevante Informationen.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

4. Die Erklärung bezieht sich konkret auf den dargestellten Fall und erklärt ihn kontextbezogen.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

5. Die Erklärung ist für Laien angemessen einfach gehalten.

1	2	3	4	5
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Die Sprache der Erklärung ist kohärent und logisch aufgebaut.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

7. Die Formulierungen sind flüssig und gut lesbar.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

8. Die Erklärung ist präzise, vermeidet unnötige Wiederholungen und Abschweifungen.

- | | | | | |
|-----------------------|-----------------------|-----------------------|----------------------------------|-----------------------|
| 1 | 2 | 3 | 4 | 5 |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> |

Offene Fragen:

- Gibt es fachliche Fehler oder kritische Auslassungen?

- Welche Aspekte der Erklärung könnten für Laien missverständlich sein?

Fachbegriffe ohne vereinfachte Umschreibung

Die Erklärung der Matrix- und Pfadsuche ist zwar korrekt, könnte aber für Nicht-Informatiker schwer vorstellbar sein, weil visuelle oder anschauliche Beispiele fehlen.

Grenzen wie „20 Einheiten“ oder „100 Sekunden“ werden nicht in ihrer praktischen Bedeutung erläutert (z. B. ob das viel oder wenig ist, und warum gerade diese Werte

Die Begriffe Diagonal, horizontal, vertikal können einen Laien verwirren bzw. man kann sich das eventuell nicht gut vorstellen

- Welche konkreten Verbesserungen schlagen Sie inhaltlich oder didaktisch vor?