

Analyse und Vorhersage des Land-Ozean-Datensatzes der NASA

1. Einleitung

In diesem Projekt möchte ich den NASA-Datensatz zu globalen Temperaturabweichungen (Anomalien) analysieren.

Dieser Datensatz enthält die Abweichungen von der mittleren Temperatur des „Referenzzeitraums 1951-1980“ für die Jahre von 1880 bis 2025.

Dabei möchte ich statistische Analysen durchführen, den Temperaturverlauf analysieren, saisonale Unterschiede untersuchen und betrachten, wie stark sich einzelne Monate auf den Jahresdurchschnitt auswirken.

Weiters möchte ich eine Prognose für die Jahresdurchschnittstemperatur bis 2050 erstellen, sowie untersuchen, wie gut sich Temperaturen mit LSTM prädictieren lassen.

2. Datengrundlage

Der **Land-Ocean Global Mean Temperature Data**" - Datensatz stammt von der Website <https://data.giss.nasa.gov/gistemp/> und enthält folgende Spalten:

- **Year (von 1880-2025):** gibt das Jahr der Messung an
- **Monate (Jan, Feb, ..., Dec):** die Temperaturabweichung (Anomalie) von der Durchschnittstemperatur im Zeitraum 1951-1980 des jeweiligen Monats
- **J-D: Jahresdurchschnittsanomalie**
- **D-N: Dezember-November-Anomalie**
- **DJF: Winter-Anomalie (Dezember, Jänner, Februar)**
- **MAM: Frühlings-Anomalie (März, April, Mai)**
- **JJA: Sommer-Anomalie (Juni, Juli, August)**
- **SON: Herbst-Anomalie (September, Oktober, November)**

3. Datenaufbereitung

- Der Datensatz wurde ohne die Überschrift eingelesen
- Die erste Zeile (Jahr 1880) wurde wegen fehlender Einträge entfernt
- Die Indizierung wurde neu gesetzt.

4. Statistische Zusammenfassung:

	J-D	DJF	MAM	JJA	SON
Anzahl	145	145	145	145	145
Mittelwert	0.083	0.078	0.083	0.07	0.099
Standardabweichung	0.403	0.431	0.422	0.388	0.401
Minimum	-0.49 (1909)	-0.68	-0.58	-0.50	-0.52
25%- Quartil	-0.2	-0.23	-0.26	-0.22	-0.19
50%- Quartil	-0.03	-0.03	0	-0.04	-0.01
75%- Quartil	0.32	0.37	0.31	0.32	0.28
Maximum	1.28 (2024)	1.36	1.28	1.23	1.41

Die Tabelle beschreibt die statistische Zusammenfassung der Spalten J-D (Jahresdurchschnitt), sowie der Spalten DJF, MAM, JJA, SON (siehe Datengrundlage).

Der Mittelwert über alle Temperaturabweichungen in den Jahren 1881 bis 2025 beträgt 0.083°C . Das bedeutet, dass die mittlere Temperatur über die Jahre 1881 bis 2025 um 0.083°C größer ist als die mittlere Temperatur in den Jahren 1951-1980.

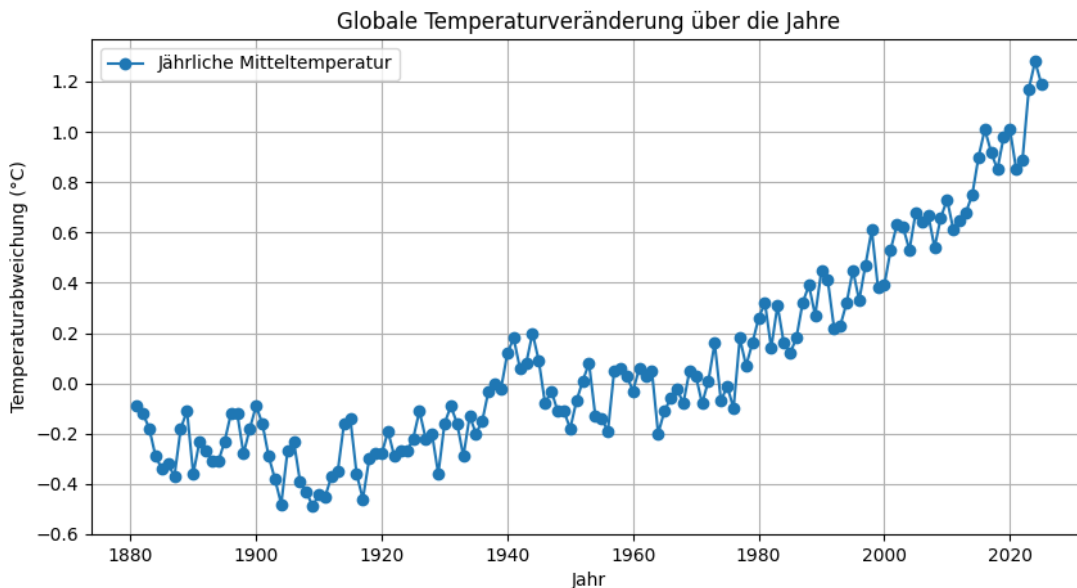
Bei Betrachtung der Jahreszeiten fällt auf, dass die gemittelte Anomalie im Herbst am größten war ($0.099^{\circ}\text{Celsius}$). Die stärksten Schwankungen bei der Temperatur gab es jedoch im Winter, die geringsten Schwankungen im Sommer.

Das kälteste Jahr war 1909 mit einer Anomalie -0.49°C , während das wärmste Jahr 2024 war mit einer Anomalie von 1.28°C .

Diese Daten geben bereits einen Einblick, wie sehr sich die Erdoberflächentemperatur seit 1880 erhöht hat. In der weiteren Analyse wird dieser Temperaturanstieg noch deutlicher sichtbar.

4. Explorative Datenanalyse

- **Globale Temperaturveränderung im Zeitraum 1880-2025**



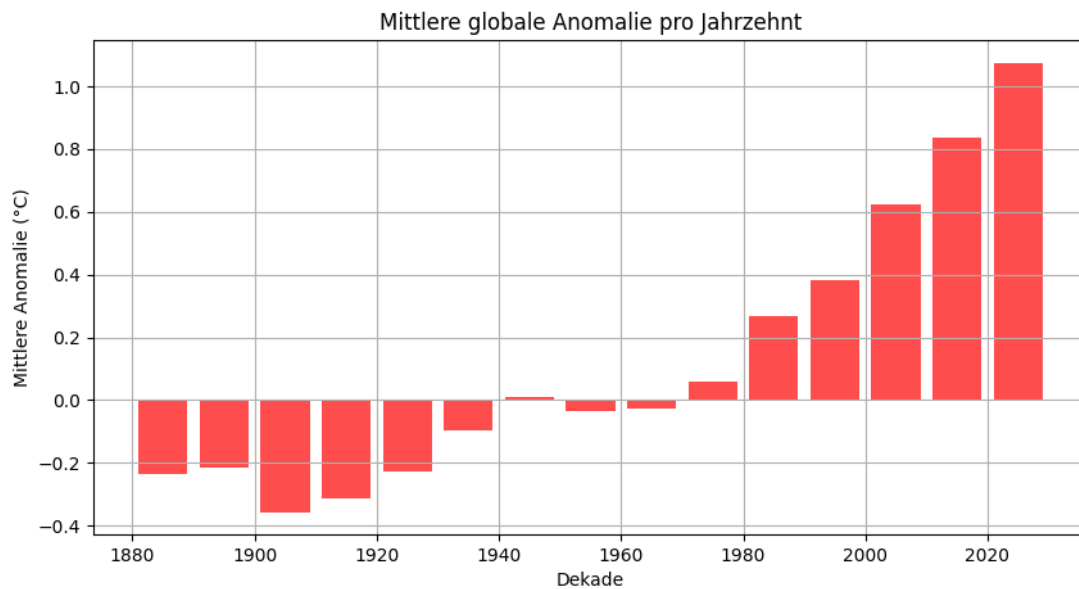
Die obere Grafik beschreibt die Entwicklung der Erdoberflächentemperatur von 1880 bis 2025. Die hier angegebene Temperatur ist die Temperaturabweichung (Anomalie) im jeweiligen Jahr von der mittleren Temperatur im Referenzzeitraum 1951-1980.

Im Zeitraum 1880 bis 1940 ist keine tendenzielle Steigung der Temperatur klar ersichtlich und die Anomalie verläuft in einem Bereich von -0.6° bis 0° Celsius. Danach erfolgte ein klar ersichtlicher Temperaturanstieg in den Jahren 1940- 1945, der zweifellos auf den Zweiten Weltkrieg zurückzuführen ist. In den Folgejahren bis 1980 blieb die Temperaturabweichung mit einigen Schwankungen stets im Bereich von -0.2° bis 0.2° Celsius.

Erst seit den 80er Jahren ist ein tendenzieller Anstieg der Erdoberflächentemperatur klar zu erkennen. In den Jahren 2000 bis 2010 schien sich die Anomalie zunächst auf einen Wert von ca. 0.5° bis 0.7° Celsius zu stabilisieren. Danach folgte jedoch erneut ein rasanter Anstieg der Temperatur und 2016 wurde erstmals die 1° - Grenze überschritten. Zunächst stabilisierte sich die Erdoberflächentemperatur in einem Bereich von 0.8° bis knapp über 1° Celsius (auch bedingt durch die Coronakrise), ehe in den Jahren 2023 und 2024 ein erneuter beunruhigender Temperaturanstieg bis auf 1.28° Celsius erfolgte (Ende der Pandemie). Zuletzt im Jahre 2025 sank die Jahresdurchschnittsanomalie wieder leicht auf 1.19° C.

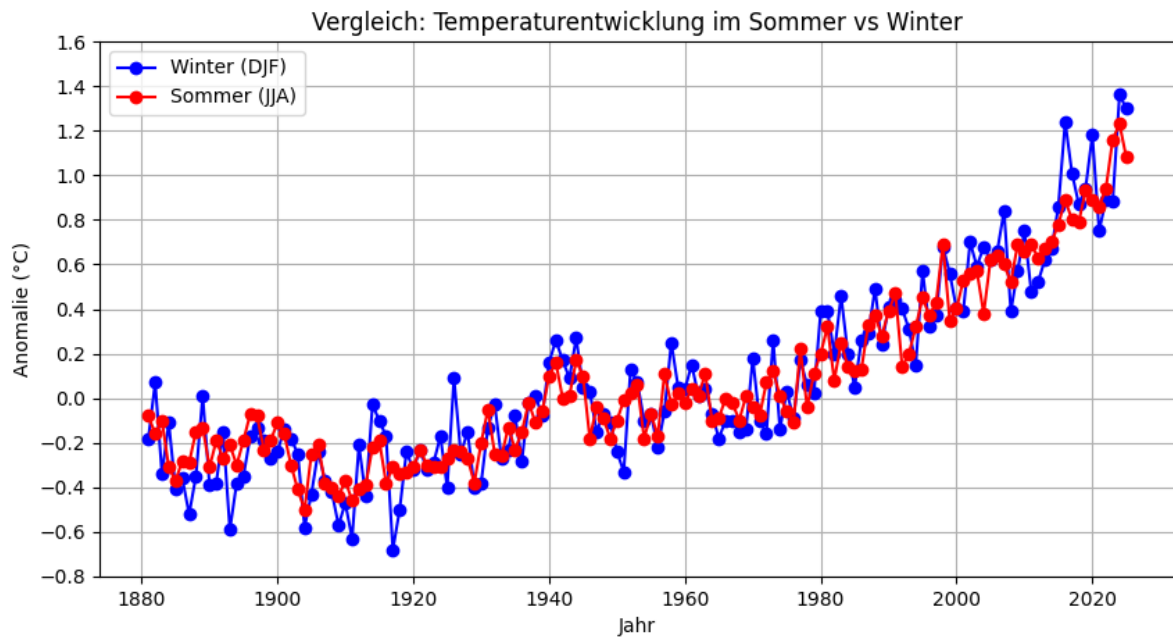
Die weitere Entwicklung in den Folgejahren ist auch deswegen relativ unklar und wird von vielen Faktoren (z.B. Klimapolitik) abhängen.

- **Durchschnittliche Anomalie pro Jahrzehnt**



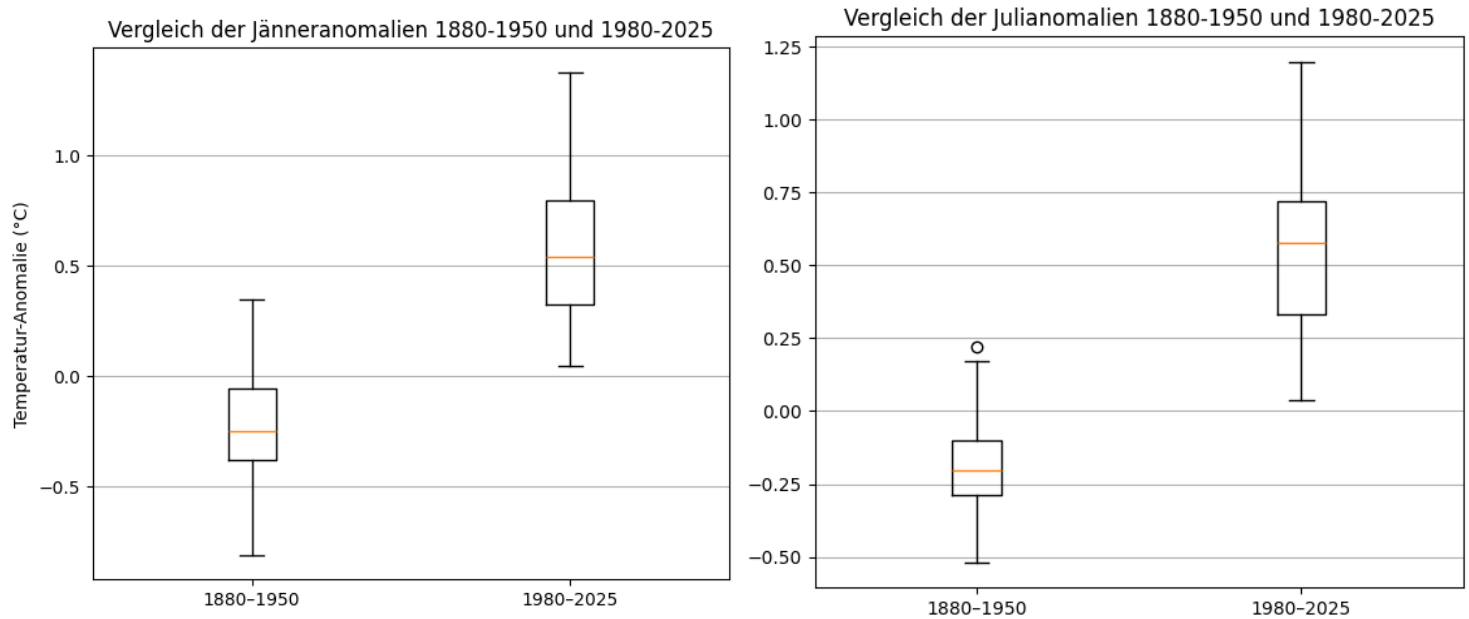
Die obere Grafik zeigt die Anomalie für jedes Jahrzehnt von 1880 bis 2024 an. Das kälteste Jahrzehnt war dabei der Zeitraum 1901-1910, danach stieg die durchschnittliche Temperatur in jedem Jahrzehnt bis 1950 um ca. 0.1° Grad an. 1950-1960 war dann das erste Jahrzehnt seit 1900-1910, das kühler als das vorige Jahrzehnt war und wie Grafik zeigt, sollte es auch das letzte bleiben. Von 1940 bis 1980 stieg die Temperatur nur unmerklich an, danach war jedes Jahrzehnt deutlich wärmer als das vorige. Seit 2000 war jedes Jahrzehnt sogar um mehr als 0.2° wärmer als das vorige. Auf den ersten Blick scheint die Erdoberflächentemperatur seit 1980 annähernd linear anzusteigen. Allerdings ist zu beachten, dass der letzte Balken nur die durchschnittliche Temperatur von 2021 bis 2025 anzeigt und dass die tatsächliche mittlere Temperatur im Jahrzehnt 2020-2030 sich daher noch leicht verändern kann.

- **Vergleich der Temperaturentwicklungen im Sommer versus Winter**



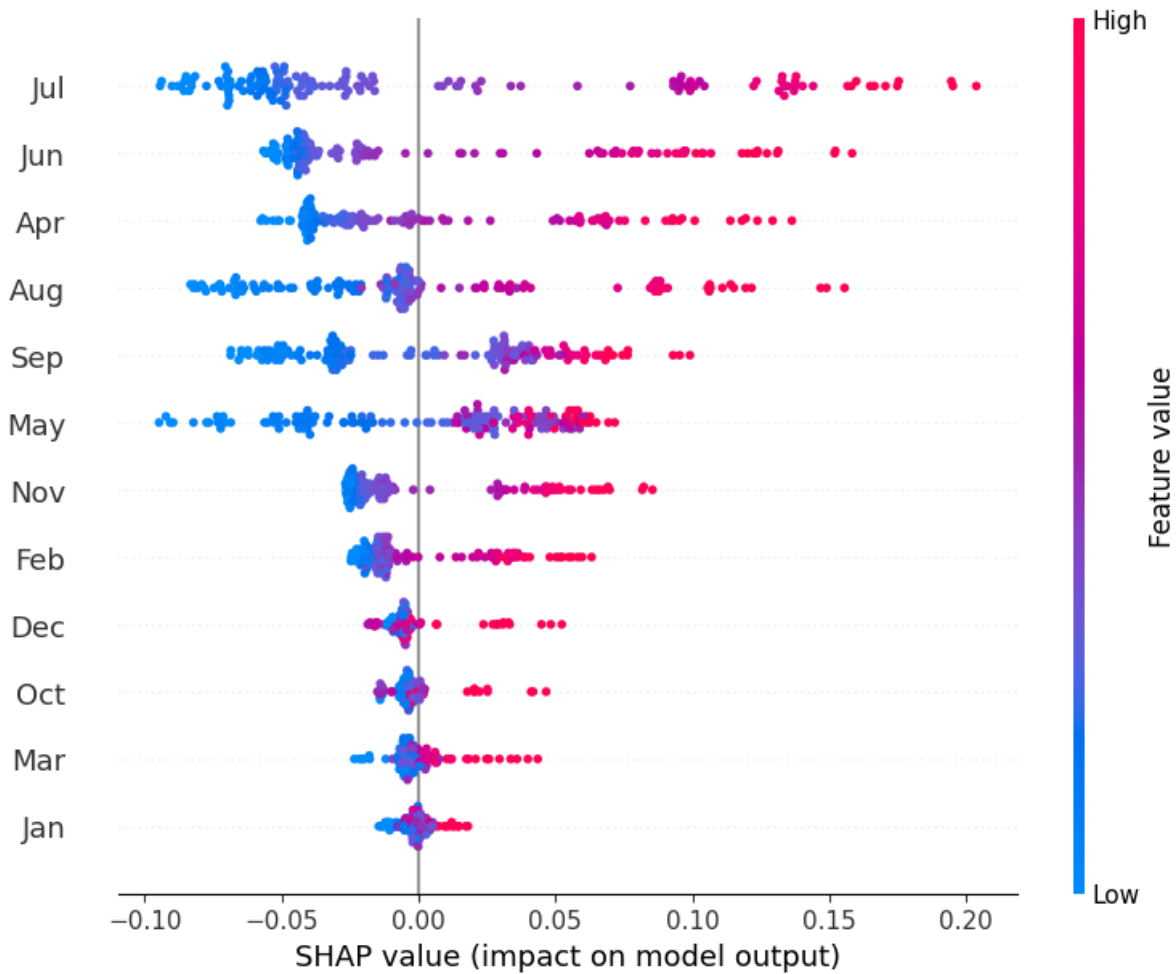
Vergleicht man die Temperaturentwicklung im Sommer und Winter, fällt auf, dass es im Zeitraum 1880- 1930 stärkere Schwankungen im Winter als im Sommer gab. Vor allem bis 1920 gab es in relativ regelmäßigen Abständen kalte Winter, die nach 1920 in dieser Form nie mehr auftraten. Im Zeitraum 1940 bis Mitte der 2010er Jahre gab es wenig Auffälligkeiten beim Sommer-Winter-Vergleich. In den letzten gab es jedoch vermehrt warme Winter und in den letzten beiden Jahren stiegen auch die Temperaturen im Sommer stark an (im Vergleich zu den Jahren davor).

- **Jänner- und Juli-Vergleich**



In der oberen Grafik werden die Jänner- und Juli-Anomalien auf dem Zeitraum 1880-1950, sowie im Zeitraum 1980-2025 betrachtet. Man erkennt sofort, dass der Abstand zwischen dem 25% und 75%-Quartil im Zeitraum 1980-2025 deutlich größer ist, was bedeutet, dass die Temperaturen im Zeitraum 1880-1950 stabiler waren. Auch sind die Whisker im Zeitraum 1980-2025 nach oben fast doppelt so lang wie im Zeitraum 1880-1950.

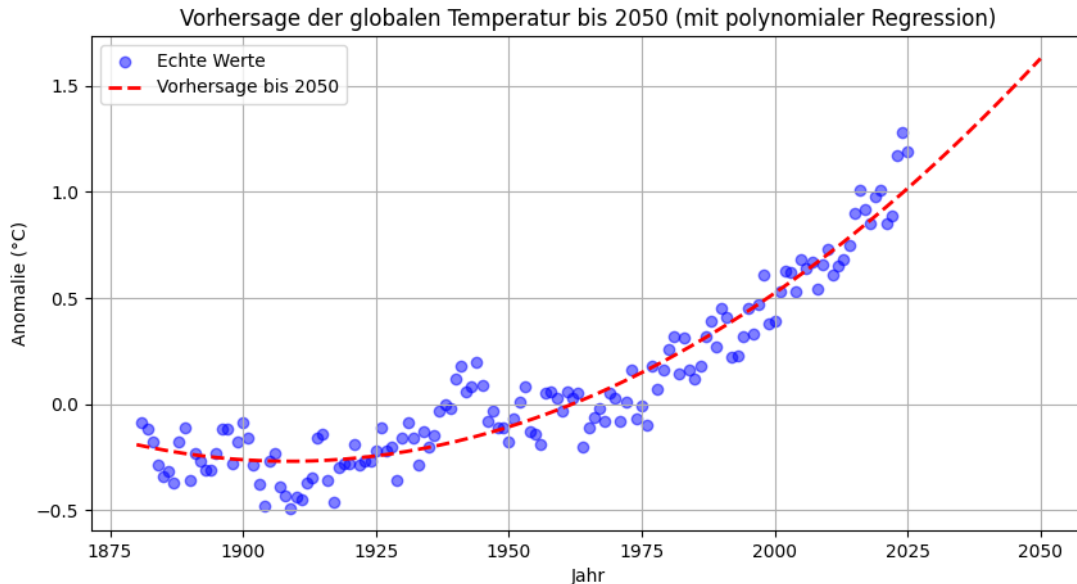
- Einfluss der Monate auf die Jahresdurchschnittsanomalie



In der oberen Grafik wird mithilfe eines Random-Forest-Regressors die Wichtigkeit eines jeden Monats auf die Jahresanomalie untersucht. Diese Wichtigkeit wird mithilfe der SHAP-values beschrieben. Diese beschreiben vereinfacht gesagt, wie viel Grad jeder Monat zur Jahresdurchschnittsanomalie beiträgt. Klar erkennbar ist hier, dass es viele Jahre gibt, wo die Monate Juni, Juli und April stark zum Jahresmittel beitragen. Die Jänneranomalien tragen hingegen nur sehr wenig zum Jahresdurchschnitt bei.

5. Prognosen bis 2050 und LSTM

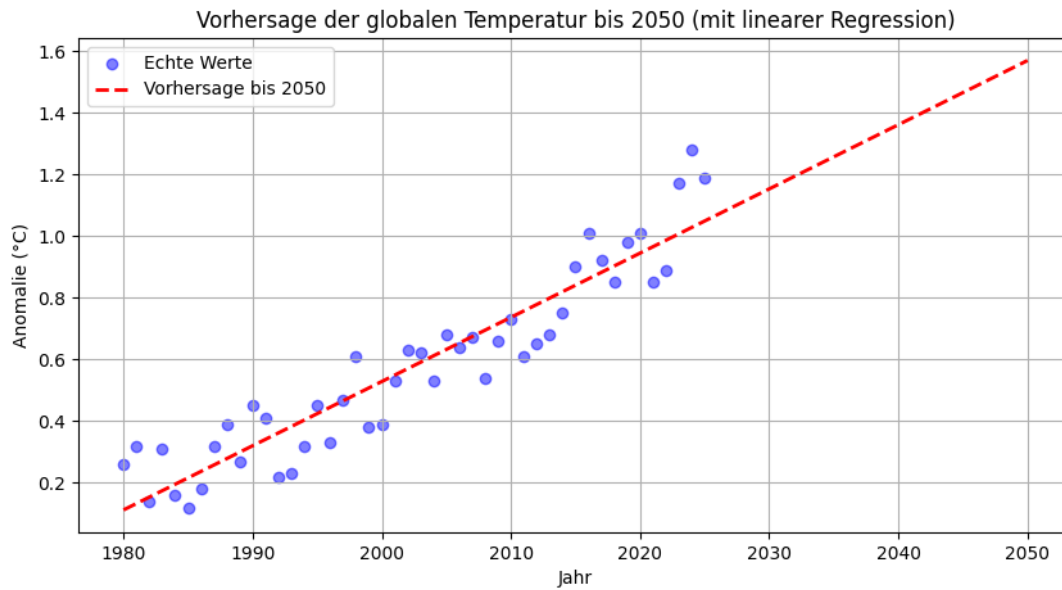
- **Prognosen bis 2050 mit polynomialer Regression**



Die obere Grafik trifft eine Vorhersage mithilfe einer polynomialen Regression mit allen Daten ab 1881. Dabei wird ein Polynom zweiten Grades gesucht, dass die Temperaturanomalien von 1881 bis 2025 am besten annähert. Basierend auf dieser Vorhersage wird die Temperaturabweichung von ca. 1.2°C auf bis ca. 1.62°C ansteigen. Das entspricht einem Temperaturanstieg der Erdoberfläche von ca. 0.4°C in 25 Jahren.

Ein Polynom höheren Grades würde keine realistische Prognose ergeben, da ein solches Polynom zu stark von dem bereits erkannten annähernd linearen Verlauf der Temperaturanomalie ab 1980 abweichen würde.

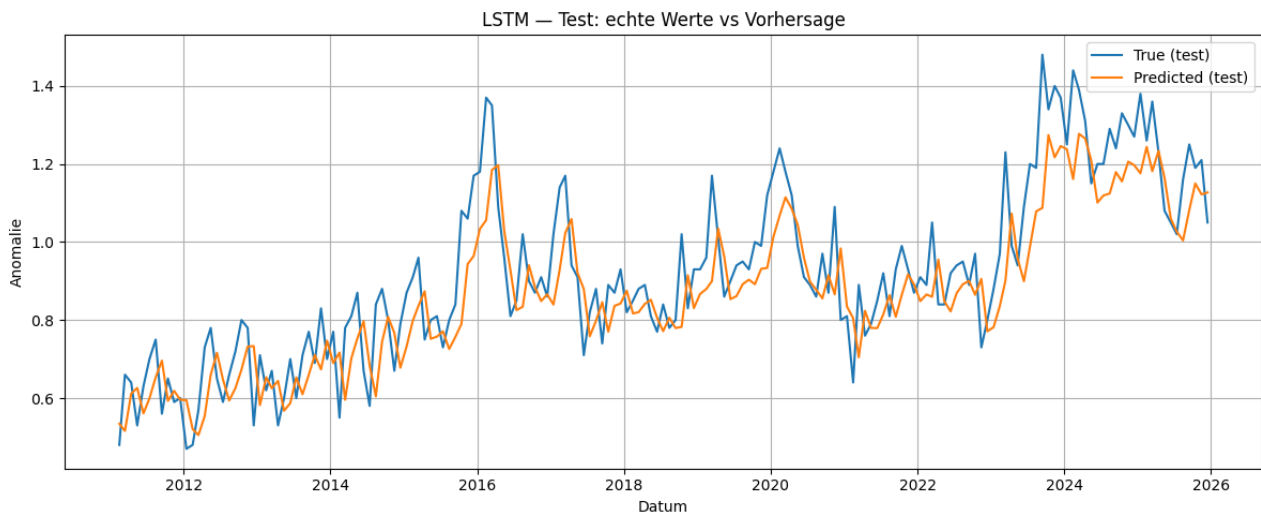
- **Prognosen bis 2050 mit linearer Regression (nur mit den Daten ab 1980)**



Die obere Grafik trifft eine Vorhersage mithilfe einer linearen Regression nur mit den Daten ab 1980. Eine lineare Regression ist hier aus zwei Gründen besser als eine polynomiale Regression: Erstens ist bereits aus der Betrachtung der durchschnittlichen Anomalie pro Jahrzehnt erkennbar, dass die Temperatur ab 1980 annähernd linear verläuft. Zweitens ist der starke Anstieg von 2022 auf 2023 coronabedingt passiert und auch aufgrund des leichten Temperaturabfalls von 2024 auf 2025 ist nicht zu erwarten, dass es weitere solche sprunghaften Anstiege geben wird. Die geschätzte Anomalie 2050 ist mit dieser Prognose in etwa 1.57°.

Beide Prognosen betrachtet beträgt die geschätzte Jahresdurchschnittsanomalie im Jahr 2050 etwa 1.6°C. (Anstieg um ca. 0.4°C von 2025)

- **Wie gut schätzt LSTM die Temperaturen von 2010-2025?**



Mithilfe eines LSTM-Netzwerks (Long Short-Term Memory) lässt sich die monatliche Temperaturanomalie für einen Monat auf Basis einer festgelegten Anzahl (= Window) vorhergehender Monate prognostizieren. Das Window in diesem Bsp wurde auf 120 Monate gesetzt. (entspricht 10 Jahre).

Das heißt, dass in der oberen Grafik die Temperaturanomalie für jeden Monat aus den Temperaturanomalien der letzten 120 Monaten berechnet wurde.

Für diese Berechnung wurde das LSTM-Modell von 1881 bis 2000 trainiert. Daher ist es auch wenig verwunderlich, dass die prognostizierten Werte durchgehend unter den tatsächlichen Werten liegen, da im Bereich 1881 bis 2000 der durchschnittliche Temperaturanstieg geringer ist als in den Jahren 2010-2025.

Der MAPE schwankt in etwa zwischen 10.5 und 12%.

Trainiert man das LSTM-Modell mit allen Daten von 1881 bis 2025, so kann man eine etwas genauere Prognose für die Jänneranomalie 2026 bestimmen. Diese beträgt in dem Fall in etwa 1.09° Celsius und liegt um ein paar Zehntelgrad unter der Jänneranomalie 2025 (1.38° C).

Zu bemerken ist auch, dass Prognosen für weitere Monate im Jahr 2026 und auch danach mit LSTM wenig Sinn machen, da dann immer mehr und mehr prognostizierte Werte für die Prognose notwendig sind, wodurch die Vorhersagen immer ungenauer und letztendlich unbrauchbar werden.

6. Fazit

Aus den Daten geht klar hervor, dass die Erdoberflächentemperatur seit 1880 auf 1.3° Celsius gestiegen ist. Von 1940 bis 1980 stieg die Temperatur nur geringfügig an, seit 1980 war jedes Jahrzehnt im Schnitt deutlich wärmer als das Jahrzehnt davor. Insbesondere in den letzten Jahren gab es sehr warme Winter und Sommer.

Beim Vergleich der Jänner- und Juli-Anomalien in den Zeiträumen 1880-1950, sowie 1980-2025 war klar erkennbar, dass die Temperaturen im Zeitraum 1980-2025 instabiler sind.

Aus einer SHAP-Analyse mithilfe eines Random-Forest-Regressors ging hervor, dass der Juli der einflussreichste Monat auf die Jahresdurchschnittsanomalie ist, während der Jänner am wenigsten Einfluss auf diesen Wert hat.

Die Prognosen basierend auf allen Daten sagen einen Temperaturanstieg von ca. 0.4° Celsius von 2025 auf 2050 voraus. Allerdings ist zu beachten, dass eine allgemeine Vorhersage sich als schwierig erweist, vor allem, weil in diese Prognose nur Temperaturdaten und keine anderen Faktoren miteinfließen, die den weiteren Temperaturverlauf doch drastisch beeinflussen können. Daher sollte diese Prognose nicht zu ernst genommen werden.

Mittels LSTM kann man sehen wie gut die Monatsanomalien von 2010 bis 2025 aus den Monatsanomalien von 1880 bis 2000 gelernt werden können. Dabei kann auch eine Vorhersage für den Jänner 2026 gemacht werden, weitere Vorhersagen sind jedoch mit LSTM problematisch.

Quelle:

GISTEMP Team, 2025: *GISS Surface Temperature Analysis (GISTEMP), version 4*. NASA Goddard Institute for Space Studies. Dataset accessed 2026-01-18 at <https://data.giss.nasa.gov/gistemp/>

Lenssen, N., G.A. Schmidt, M. Hendrickson, P. Jacobs, M. Menne, and R. Ruedy, 2024: [A GISTEMPv4 observational uncertainty ensemble](#). *J. Geophys. Res. Atmos.*, **129**, no. 17, e2023JD040179, doi:10.1029/2023JD040179.