**Kurze Erklärung der vier Modelle und ihre Auswahl:**

*Titel: Auswahl von ML-Modellen für die Einkommensklassifikation*

Im Rahmen der Einkommensklassifikation haben wir uns für die Anwendung von vier Machine Learning (ML)-Modellen entschieden: K Nearest Neighbor (kNN), Random Forest, Gradient Boosted Trees und Naive Bayes. Jedes Modell bietet spezifische Vorzüge, die auf die Anforderungen unserer Aufgabe zugeschnitten sind.

1. **K Nearest Neighbor (kNN):**
   * *Beschreibung:* KNN ist ein einfacher, aber leistungsfähiger Algorithmus, der auf der Idee basiert, dass ähnliche Datensätze in der Regel die gleiche Klassifikation haben. Er eignet sich besonders für die Identifizierung von Mustern in den Daten.
   * *Warum kNN:* Aufgrund der Annahme, dass ähnliche Einkommensklassen nahe beieinander liegen, kann kNN gut geeignet sein, um lokale Muster zu erkennen und präzise Vorhersagen zu machen.
2. **Random Forest:**
   * *Beschreibung:* Random Forest ist ein Ensemble-Lernalgorithmus, der auf Entscheidungsbäumen basiert. Er erstellt mehrere Bäume und kombiniert ihre Vorhersagen, was zu einer robusten und genauen Klassifikation führt.
   * *Warum Random Forest:* Random Forest eignet sich gut für komplexe Datenstrukturen und bietet eine hohe Genauigkeit. Durch die Kombination mehrerer Bäume wird Overfitting reduziert.
3. **Gradient Boosted Trees:**
   * *Beschreibung:* Gradient Boosted Trees ist ein weiterer Ensemble-Lernalgorithmus, der Entscheidungsbäume sequenziell erstellt. Jeder Baum korrigiert die Fehler des vorherigen Baumes, was zu einem starken Modell führt.
   * *Warum Gradient Boosted Trees:* Dieser Algorithmus ist effektiv bei der Handhabung komplexer Beziehungen in den Daten. Er bietet eine hohe Genauigkeit und ist besonders nützlich, wenn präzise Vorhersagen erforderlich sind.
4. **Naive Bayes:**
   * *Beschreibung:* Naive Bayes basiert auf dem Bayes'schen Theorem und geht von der naiven Annahme aus, dass Merkmale unabhängig voneinander sind. Dieser Ansatz eignet sich gut für einfache Klassifizierungsaufgaben.
   * *Warum Naive Bayes:* Naive Bayes ist schnell zu implementieren und funktioniert gut mit großen Datensätzen. Es ist besonders effizient, wenn die Unabhängigkeitsannahme für die Merkmale erfüllt ist.

**Entscheidungsgrundlage:** Die Auswahl der Modelle erfolgte basierend auf der Komplexität der Daten, der gewünschten Genauigkeit und der Interpretierbarkeit der Ergebnisse. KNN wurde gewählt, um lokale Muster zu erfassen, Random Forest für komplexe Strukturen, Gradient Boosted Trees für präzise Vorhersagen und Naive Bayes für schnelle und effiziente Klassifizierung.

Insgesamt bieten diese Modelle eine ausgewogene und leistungsfähige Basis für die Einkommensklassifikation, wobei jedes Modell seine Stärken je nach Datenstruktur und Anforderungen der Aufgabe ausspielt.