

# Projektarbeit 1 – Titanic



**Bearbeitungszeit:**

15.05.24 – 03.06.24

Gruppenarbeit: max. 4 Mitglieder/Gruppe.

**Abgabe:** C++ vorab in ILIAS. In den Übungen am 03.06.24, werden Ihre Programme vor Ort geprüft & testiert.

Das **Bestehen** der **beiden Projekte** ist **notwendige** und hinreichende Bedingung, um die Übungen zur Vorlesung Info2 zu bestehen.

**Aufgabe:** Nutzen Sie maschinelles Lernen um Vorherzusagen, ob eine Person den Titanic Untergang wahrscheinlich überlebt hätte. Schreiben Sie hierzu ein C++ Programm, welches

- die Daten aus der Datei titanic.csv in eine geeignete Datenstruktur lädt.
- einen Entscheidungsbaum mit dem CART-Algorithmus auf Basis der Daten erstellt (Training).
  - Nutzen Sie mindestens 3 verschiedene Features/Eigenschaften (z.B. Alter, Geschlecht, Klasse, Mitreisende -, Angehörige vorhanden, etc.), wobei mindestens eine Eigenschaft ein numerischer Wert sein muss (z.B. Alter oder Kosten des Tickets).
  - Als Label/Marker dient die Angabe, ob der Reisende überlebt hat.
  - Die maximale Tiefe des Baumes, soll vom Nutzer gewählt werden können.
- auf Basis des trainierten Entscheidungsbaumes eine Vorhersage treffen kann (Inferenz), ob eine Person den Untergang der Titanic überlebt hätte.

Hierbei ist eine CSV Datei wie folgt strukturiert:

```
1 Survived,Pclass,Name,Sex,Age,Siblings/Spouses Aboard,Parents/Children Aboard,Fare
2 0,3,Mr. Owen Harris Braund,male,22,1,0,7.25
3 1,1,Mrs. John Bradley (Florence Briggs Thayer) Cumings,female,38,1,0,71.2833
4 1,3,Miss. Laina Heikkinen,female,26,0,0,7.925
5 1,1,Mrs. Jacques Heath (Lily May Peel) Futrelle,female,35,1,0,53.1
6 0,3,Mr. William Henry Allen,male,35,0,0,8.05
```

Nach der Spaltendefinition (Zeile 1) folgen die Samples/Beispiele zeilenweise. Hier ist z.B. in der zweiten Zeile zu entnehmen, dass der Mann Mr. Braund mit 22 Jahren ohne Kinder/Eltern aber mit einem Angehörigen ertrunken ist und für die Fahrt in der 3ten Klasse 7.25£ zahlte.

Testen Sie Ihren Entscheidungsbaum (Vorhersage: Überlebt ja/nein) mit dem Beispielen:

?,3,Mr. Jack Dawson,male,23,0,0,6.14

?,1,Mr. Donald Trump,male,78,1,0,120.8

?,1,Miss. Rose DeWitt Bukater,female,17,0,1,78.1

Fügen Sie bitte folgenden Kommentar in den Source-Code Ihre Projektarbeit vor der Abgabe via Ilias ein:

/\*

Projektarbeit 1 – Titanic

Gruppe:

[author #1]

[author #2]

[author #3]

[author #5]

Wir stimmen der Veröffentlichung unseres Source Code in anonymisierter Form zu.

oder

Wir stimmen der Veröffentlichung unseres Source Code mit Namensnennung zu.

oder

Wir stimmen der Veröffentlichung unseres Source Code **\*nicht\*** zu.

Copyright (C) [year] [authors #1-#4]

SPDX-License-Identifier: MIT

\*/

Zusatzaufgabe (**optional**):

Testen Sie die Genauigkeit (Anzahl korrekter Vorhersagen/ Anzahl aller) Ihrer Vorhersage (auf dem Trainingsset) als Funktion der maximalen Baumtiefe.

Was stellen Sie fest? Wie können Sie dieses Verhalten erklären?