

Project proposal: Titanic – Machine Learning from Disaster

Authors: Katarzyna Bocian, Kirill Korzyuk, Adam Kozakowski

In our project we plan to study Titanic problem from historical data and get familiar with Kaggle competition structure. In to do so we will examine updated datasets from Kaggle¹ and try to complete them with data found online. Additionally, as a beginners in ML, we want to make use of Kaggle tutorials provided to this project. ²

Motivation:

This project will be about learning from mistakes of the past – Titanic disaster. We will study the historical data about survivors to conclude, what were the main characteristics of passengers that prevailed on their lives. Results of study could be used to design next voyages to be more safe and better risk assessed.

Method:

Understanding database and doing necessary adjustments (exploratory data analysis). Defining machine learning model in Python (simple perceptron, PCA analysis, support vector machine SVM, regression model).

Intended experiments:

Finding most important factor of survival. Finding factors with little impact. Adjusting hyperparameters. Feature engineering. Comparing machine learning algorithms and trying to combine them (ensemble learning).

¹ Kaggle Titanic dataset: <https://www.kaggle.com/competitions/titanic/data>

² Kaggle Titanic Tutorial: <https://www.kaggle.com/code/alexisbcook/titanic-tutorial/notebook>