


Helge Bjorland

An Interactive Data Science Tutorial


last run 6 months ago · Python notebook · 29275 views

using data from [Titanic: Machine Learning from Disaster](#) ·  Public

▲

120

voters



Notebook

Code

Data (1)

Output (1)

Comments (36)

Log

Versions (14)

Forks (1715)

Fork Notebook

Notebook

An Interactive Data Science Tutorial

Based on the Titanic competition on Kaggle (<https://www.kaggle.com/c/titanic>)

by Helge Bjorland & Stian Eide

January 2017

Content

1. Business Understanding (5 min)
- Objective

• Description
2. Data Understanding (15 min)
- Import Libraries

• Load data

• Statistical summaries and visualisations

• Excersises
3. Data Preparation (5 min)
- Missing values imputation

• Feature Engineering
4. Modeling (5 min)
- Build the model
5. Evaluation (25 min)
- Model performance

• Feature importance

• Who gets the best performing model?
6. Deployment (5 min)
- Submit result to Kaggle leaderboard

Adopted from Cross Industry Standard Process for Data Mining (CRISP-DM) (<http://www.sv-europe.com/crisp-dm-methodology/>)



1. Business Understanding

1.1 Objective

Predict survival on the Titanic

1.2 Description

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic

