**ENE4062-01 Fall 2023  
Machine Learning Theory Project Proposal**

**Prediction of Fares Paid by Titanic Passengers**

조윤지, 강민지, Titouan MILLET

**Division of Electronics & Electrical Engineering, Dongguk University, Seoul, Republic of Korea**

**Description**

The Titanic, a legendary ocean liner, tragically became famous for its ill-fated maiden voyage in 1912. What was initially a dream of grandeur turned into a nightmare as the ship collided with an iceberg, resulting in its tragic sinking. The Titanic's sinking, one of the most iconic maritime disasters in history, left an indelible mark on the collective memory and prompted significant reforms in maritime safety.

In this fascinating historical context, a compelling question arises: Can we predict the fares paid by Titanic passengers based on certain personal characteristics? This question gains relevance as it allows us to explore the factors that influenced passengers' financial decisions aboard the Titanic. By applying linear regression, this project aims to establish a model capable of estimating fares based on these characteristics and to analyze how each factor may have played a role.

Linear regression is the appropriate analytical method for addressing this problem, as it is well-suited for modeling linear relationships between a dependent variable (fare paid) and multiple independent variables (passenger characteristics). In this case, we seek to quantify how these characteristics influence the fare. Linear regression enables us to construct a predictive model, considering the effect of each characteristic in a linear manner, which is suitable for estimating fares paid.

Key Steps for Resolution:

1. An initial and crucial data preparation phase will involve cleaning the Titanic dataset to ensure the reliability of the results.
2. Thoughtful consideration will guide the selection of features to include in the model to accurately identify those influencing the fare paid.
3. Implementation of linear regression, our prediction tool, will be closely monitored to develop a robust model.
4. To assess the model's relevance, metrics such as the coefficient of determination R² and root mean square error will be used.
5. Lastly, interpretation of the model's coefficients will unveil the relationships between characteristics and fares, providing an in-depth understanding of these financial interactions.

Keywords: Titanic, Fare Prediction, Linear Regression, Machine learning, Data Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Project schedule** | | | | |
|  | **November**  **1-9** | **November**  **9-30** | **Décembre**  **1-7** | **Décembre**  **8-14** |
| **Writing Proposal** |  |  |  |  |
| **Code** |  |  |  |  |
| **Project presentation PowerPoint** |  |  |  |  |
| **Project report** |  |  |  |  |