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

**Prediction of Survival Probability of Titanic Passengers**

Yunji Jo, Minji Kang, Titouan MILLET

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

**Description**

The Titanic, a magnificent ocean liner, left an indelible mark on history through its tragic sinking in 1912. On its maiden voyage, a dream of grandeur turned into a nightmare as the ship struck an iceberg and descended into the frigid waters of the Atlantic. The Titanic's sinking remains one of the most iconic maritime disasters, claiming many lives. This event triggered significant reforms in maritime safety and left an enduring impact on history.

Within the Titanic's historical context, a crucial question emerges: Can we predict the probability of a passenger's survival based on their personal characteristics? This question is of great significance as it allows us to explore how factors such as ticket class, gender, age, and more influence the likelihood of survival.

The project aims to understand the dynamics behind the critical decisions aboard the Titanic and to predict the probability of survival. We intend to utilize several machine learning models in the analysis of Titanic datasets. The models we will use are logistic regression as linear models and DT and svm as nonlinear models. Each of these models has these characteristics. And each of these models has these features.

SVM(Support Vector Machine): is a supervised learning algorithm used for classification and regression problems. It primarily maps data into a high-dimensional space and seeks to find the optimal decision boundary (separating hyperplane). It classifies data by using support vectors (data points closest to the decision boundary) to achieve the best separation. It is applicable to both linear and non-linear problems, and it can utilize kernel methods to find non-linear patterns.

Decision Trees (DT) are a supervised learning algorithm used for both classification and regression tasks. They divide data based on feature attributes, using yes/no questions at each node of the tree. Decision trees provide a simple and intuitive model for data partitioning and prediction. They can be pruned and have feature selection capabilities to prevent overfitting.

Linear regression is a supervised learning algorithm used for regression problems. It is primarily employed to predict continuous output values. Linear regression models the linear relationship between data features and weights. The goal is to minimize the residual (error) between the predicted linear model and the actual results.

We will compare and analyze each performance of these models to find the optimal model that improves the accuracy of the prediction. This paper will provide an in-depth understanding of the factors that influenced the survival of Titanic through optimal predictive models.

Keywords: Titanic, Survival Prediction, Logistic Regression, DT, SVM. Machine learning, Data Analysis

**References**

[1] Cho Tae-ho. "Deep learning for Everyone. Third edition". Gilbott. 2022

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