

## BLM4800 Introduction to Data Mining - Project Proposal Form

Project work can be done individually or as a team work. Groups can be up to 3 people. As the number of group work increases, the scope of the project is expected to expand.

The following information about your project proposal should be filled in and a copy of this file should be uploaded to the Online.yildiz.edu.tr system. Only one student from the group needs to upload the Project Proposal form. Due date of the Project Proposal: Sunday, April 16, 2023.

**Name of Team member(s):** Parsa Kazerooni - 19011915, Özkan Özeşme - 19011009

**Name of Proposed Project:** Salary Prediction using Data Mining Techniques

**Purpose of the Project (1 paragraph):** The purpose of this project is to develop a data mining model that can predict the salary of American citizens based on various demographic and socioeconomic factors. The dataset contains information on age, education, work experience, occupation, marital status, race, and gender, which can be used to identify patterns and relationships that affect an individual's income.

Dataset: Becker, Barry & Kohavi, Ronny. (1996). Adult. UCI Machine Learning Repository.

<https://doi.org/10.24432/C5XW20>

**Project Steps and potential Methods to be Used (paragraph 1-2):** Data cleaning and preprocessing, exploratory data analysis, feature engineering, model selection and training, model evaluation and validation.

Data visualization: histograms, scatter plots, and box plots

Feature engineering: correlation analysis, polynomial features and PCA

Machine learning algorithms: decision tree, random forest, linear regression, support vector machine and neural networks