MIDTERM 1371 ML - Dataset

Department of Science, Technology, Engineering & Math, Houston Community College

ITAI - 1371 Introduction to Machine Learning

Professor Vishwa Rao

October 27th, 2025.

Dear Professor,

We would like to submit the following dataset for our midterm project:

https://www.kaggle.com/datasets/arashnic/hr-analytics-job-change-of-data-scientists

By using this dataset, we will as a team practice our learnings from the L06 data processing. We will predict the target variable in aug_train.csv, which is included in the dataset. In this dataset, target = 1 means the candidate is looking for a job change, and target = 0 means not.

We will train a supervised binary classification model on aug_train.csv using the provided features (city, city_development_index, gender, relevent_experience, enrolled_university, education_level, major_discipline, experience, company_size, company_type, last_new_job, training_hours) while treating enrollee_id as an identifier only. The companion aug_test.csv is the unlabeled test split.

Thank you,

Martin Demel, Jiri Musil, Benso Munachimso, Jasmine Campos,