



Milestone 2 Tasks

Project 1, 6, 8 Tasks

1. Preprocessing

- 1) Convert categorical features into numerical
- 2) Data normalization/scaling

2. Feature selection

- 1) Features importance evaluation
- 2) Features selection methods

3. ML Models

- 1) Select 3 ML models
- 2) Split training, validating and testing sets
- 3) Compare 3 models' performances with features selection

Project 2 Tasks

1. Preprocessing (2 approaches)

- 1) Text embedding: use one of the methods in MS 1 to embed the "text" column
- 2) Feature extraction: develop numerical features from the "text" column

2. ML Models

- 1) Research on BERT pre-trained NLP model
- 2) Select 1 ML model other than BERT to evaluate based on the developed numerical features
- 3) Compare the performances of two approaches

Project 4 Tasks

1. Preprocessing

- 1) Data normalization/scaling

2. ML Models

- 1) Select 3 ML models
- 2) Split training, validating and testing sets
- 3) Compare 3 models' performances with features selection