Determining the job posted on LinkedIn whether its real or fake.

Features:

- job title
- office location
- department
- salary offered for the job
- company information
- job description
- job requirements
- benefits
- telecommuting
- company logo exists?
- has_questions?
- Employment type
- Experience required
- Education required
- Industry
- Function

Label:

- Fake? (Whether the job posted is fake or real, if yes then the job posted fake otherwise then the job posted is real).

Delivery time: In practical week.

Requirements:

I. Text cleaning (e.g., stop words removal, removing special characters and tags using regular expressions)
II. Applying text normalization techniques (e.g., stemming and lemmatization)
III. Feature extraction using at least two techniques from the explained ones (e.g., bag of words, n-grams, tf-idf vectorization).
IV. Modeling: using at least 3 different machine learning or deep learning algorithms for classification purposes
V. Evaluation: assessing the model performance on the given test set using different evaluation metrics including accuracy, precision, recall, f1-score and confusion matrix)

Report:

- You must explain in details the preprocessing techniques you applied on the dataset
- You must explain what models did you used for classification purposes.
- You must explain the impact of using different hyperparameters in preprocessing and modeling (e.g., the hyperparameter n in n-grams technique, the hyperparameters used in each model, etc.).
- You must include the results in terms of the evaluation metrics for each model used.

Bonus:

- Applying suitable data cleaning techniques.
- Trying other feature extraction techniques than those explained ones.
- Trying advanced modeling techniques such as sequential models.
- Excellent report and analytics.

Additional notes:

- You will find in the dataset some text columns such as company information, job description, etc. you can combine them together and treat them as a single text column.
- A portion of the project grade will be based on individual evaluation. Thus, each team member should understand the project.