

Task Description ML

Util

June 2018

1 Description

In the email you received with this document, you should find an excel file 'Data_ML.xlsx' - if not please reach out to me at abdel@util.co and let me know.

The excel sheet is a list of public companies, along with a description of their business, their industry sector as defined by Thomson Reuters, their ticker symbol and their name.

The columns are:

- **Identifier** - a code that identifies a company
- **Company Name** - the company's public name
- **Business Description** - a description of the companies main activities
- **TRBC Economic Sector Name** - A classification of the company's main sector as defined by Thomson Reuters

The goal of this task is to predict a company's TRBC Economic Sector Name from the Business Description. The main objective of this task is to showcase your experience with NLP, and not necessarily your classification accuracy score, however it is required that you report your classification accuracy. You will be evaluated on the quality and readability of your code, how you set up the problem and explore the data, your ability to develop a high-performance solution efficiently (given two solutions with the same accuracy, the more efficient one is preferred) and your ability to devise creative solutions to either increase accuracy or reduce model complexity.

The only requirement is you use Python in a Jupyter Notebook for the task. The remaining libraries you choose to use are up to you, however, please ensure you specify all the dependencies/libraries that need to be installed for your notebook to run. If you feel like using Python would significantly harm your ability to showcase your best effort, let me know.