**Integrated Team Project Form**

**InfoComm Technology**

**May 2020 – July 2020**

**Title of the Project**

Self learning Named-Entity Recognition Platform

**Company Details**

Element Materials Technology Singapore

106 Tuas South Ave 2, Singapore 637159

**Industry Supervisor**

Rek Chongrek.chong@element.com

**Negotiator(s)**

Muhamed Fauzi Bin Abbas

**Academic Supervisor 1:**

Muhamed Fauzi Bin Abbas

**Academic Supervisor 2 (if any):**

Dan

**Description of the Project**

Data is often loosely entered into spreadsheets as an easy method to capture information in many parts of the business. These data in most cases are unstructured text containing spelling mistakes, short form writing, comments and more importantly key information such as product, service, people etc.

This project requires the students to build an online website and develop a machine learning model using Named-Entity Recognition (NER) techniques to extract information that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, product codes etc.

The online website should develop the following components:

1. A web application as a front end to facilitate the spreadsheet upload, display progress of the data cleansing actions taken in the backend, validate the cleansing rules from the user and display output of the model.
2. A database for storing the text identified by the machine learning model
3. A report to track the performance of the model

**Keywords for the Project**

Machine Learning, Named-entity Recognition, Web Application, Database, Exploratory Data Analysis, Data Cleansing Techniques

**Prerequisites for the Project**

Students must be familiar with at least one programming language (html5, python, database, javascript, etc.), able to code for online platform and good understanding of descriptive statistics for numerical data.

**Final Deliverable(s) for the Project**

A working online platform to facilitate the data cleansing actions and a ML model to classify the unstructured text into pre-defined categories

**Is the company willing to provide space for the students to work in their premises?**

**Yes  No**

**Additional Information**

We anticipate that this project will require not more than 4-5 students to complete the project.