Massachusetts Materials Technologies LLC
167 Prospect St Unit #4
Waltham, MA 02453
617-502-5636
HSD@ByMMT.com

ByMMT.com

11 November 2020

**Subject:** Data Processing Challenge

This challenge is created to evaluate applicants on their ability to conduct basic statistical analysis, develop a predictive model, and to present your processes and results. You are provided with an Excel file containing a~~of~~ data set of experimental data collected 50 steel samples.

**Columns:**

**Sample Name:** A unique name for each tested material.

**Seam Type:** Welding method used to produce that pipe sample.

- **ERW:** Electrical resistant welding, using no additional filler materials and joining the material by heating the target surfaces and pushing them against one another.
- **SAW:** Arc welding using filler metal, more frequent in larger pipes
- **SMLS:** Abbreviation for seamless. No welding used; pipe was directly extruded to a hollow cylinder from casted steel.

**Tensile Yield:** Destructive lab test yield strength output. This is what we want to estimate using the combination of nondestructive measurements

**NDE Yield:** Our Non-destructive tester's yield strength output on the outer surface of the pipe joint

**C to V:** Weight percentage of corresponding chemical element in the sample

**Grain Size:** Average size of the crystal grains for each sample in µm.

**Requirements:**

1. Develop a code to **read** and **process** the data.
2. Develop a machine learning model that can predict **Tensile Yield** for new materials based on your selection of features form the data given.
3. Evaluate your model using your selection of performance metrics.
4. Prepare a presentation on the algorithm you used, your results, and your evaluation with justification on the decisions you made (no longer than 15 minutes). We would also like to review your source code.

**What we are looking for:**

a) Your **programming** proficiency and understanding of data **analytics, ML, and visualization**.
b) Your **approach** to a set of data and how you can find **meaningful results** from a **raw** data set.
c) Your ability to **communicate** your results to your peers and/or supervisors.

Note: There are no right or wrong answers we only focus on your method/approach. We hope that you will not spend too much time on this exercise (1-2 hours is expected)