PIDX | **ML PROJECT PROPOSAL**

**Back End**

current table

Testing Data:

25% of current table in memory

* flag with new column test yes/no

**Learning Data:**

75% of current table in memory

**Methodology:**

Scikit Learn Library classification

1. Decision tree
2. SVM

**Utilization:**

Clarifying fields necessary & using testing data to check if correct answer is generated

**Front End**

**Current Webpage**

**Additions: (similar style as first form)**

Form entry to input for code generation – what columns are most important to classify?

Required Elements (2):

* Google Cloud
* Python Pandas
* SQL Database
* HTML/CSS/Bootstrap
* **Required: Scikit-Learn**

Resources:

* <https://towardsdatascience.com/multi-class-text-classification-with-scikit-learn-12f1e60e0a9f>
* <https://stackabuse.com/overview-of-classification-methods-in-python-with-scikit-learn/>
* <https://scikit-learn.org/stable/supervised_learning.html>
* <https://techblog.commercetools.com/boosting-product-categorization-with-machine-learning-ad4dbd30b0e8>