Additional TeknoVe Info for Project Step 3

UC 1

One of the beauties of the quality control use case is that it relies on a single technology, a visual classifier. There are several types of defects that must be looked for, however, and the team has considered several different approaches. The first approach is to train a single classifier that determines whether an item is defective or not defective, and also isolates the point of the defection using a bright pink box superimposed on the image presented. Manual inspection can then determine the type of error. Another approach the team has considered is training fifty different classifiers for the fifty most common types of defects. This would allow for the defect to be identified immediately, and may prevent unnecessary line shutdowns in factories. The team's key concern is that using the first approach with only having two categories, certain-defects or not-certain-defects, would result in a very low number of certain-defects. But creating the system with multiple categories may be untenable. Still, the single classifier approach seems like only an incremental improvement given the amount of downtime it might cause on the line.

UC 2

The tax use case is almost entirely predicated upon the use of various Natural Language Processing techniques. First, the system will ingest the entirety of each tax code. Next, it will use a natural language classifier to identify distinct segments of the tax code. After that, it will use natural language understanding to interpret each segment of the tax code and identify any connections. Lastly, it will use a series of generic predictors on top of TeknoVe's existing Enterprise Resource Planning (ERP) system to identify where elements of the tax code may best be applied and formulate an optimal tax strategy. While the team recognizes there are a lot of models to train here, they are confident in each underlying model.

UC 4

The check-engine use case is fairly simple in principle. Using hundreds of sensors, the team hopes to use a generic classifier to predict whether changes in the readings of these sensors are likely to trigger a failure in the next two weeks or not. If the model projects 80% confidence or higher in a failure, the light will be deployed. For the purposes of explainability, the team plans to run a parallel regression against each of the sensors being used to predict how the model would perform if that sensor's data was unavailable. In theory, important sensors would drop confidence back below 80% for a failure event if removed. While this approach should work in theory, some members of the team have suggested doubt about this feature and question the purpose of ML in the first place if it is not possible to create a model that is interpretable.