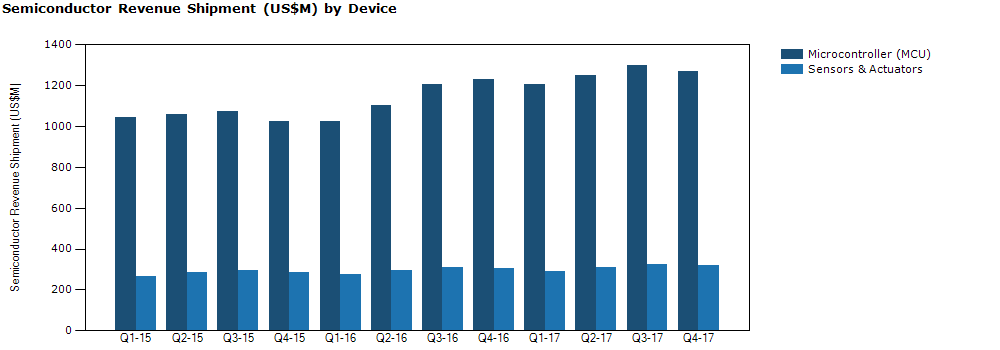
Anomaly Detection

Project Initiation Document

# Business Case Summary

**Application Space:** Smart sensors for IoT monitoring applications

IHS projects continued MCU consumption revenue growth in industrial applications in 2H2016 into 2017 but flat revenue for sensors (chart below) over the same period. The flat projection suggests either sensors are not being included in the growing industrial applications, declining value of sensors or both.



GE calls industrial IoT “the internet of *expensive* things.” It therefore follows that the growth of industrial applications should have spurred significant growth in machine-health monitoring for these expensive things. Because sending data is expensive, the ability to identify abnormal, and therefore worthwhile, data from continuous monitoring results can be a key enabler to make more sensor data useful for IIoT applications.

Machine condition monitors have successfully used vibration signatures to detect abnormal behavior for mission critical industrial and medical equipment. To make vibration signature more useful for more equipment installations, it is necessary to automate the process of training a monitor to differentiate abnormality.

In recent years:

* MCU costs have fallen
* Sensor costs have fallen
* Machine learning techniques have gone mainstream

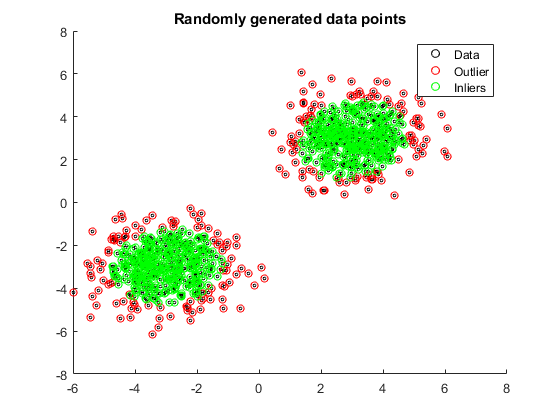
The intersection of the three trends has the potential to open a new market in low to mid-tier products for machine condition monitoring. Possibilities include pool pump that won’t burn out when the intake is dry, water heater that would shut off on its own when there is a leak, hard disk array that alerts IT when a ball bearing on a fan is going bad before the fan’s vibration causes reduced data bandwidth for the array, etc. It should be noted that vibration monitors could be suitable for abnormality detection even when the normal output would be monitored by a different mechanism, i.e. pressure sensors.

Imagine the business potential if we could convince motor manufacturers to include monitoring devices on even a fraction of low and mid-tier motors.

# High Level Project Description

This project represents a continuation of learning efforts over the last couple of years that have targeted this application space. The goal is creation of value added IP targeted at anomaly detection, specifically for machine condition monitoring.

The figure below illustrates a single-class support vector machine applied to a bimodal set of data points. The algorithm maps these data points into a higher dimensional space for the analysis.



Possible implementation choices are discussed in the Risks section below.

## Risks

This project should be considered an NTI as it contains a number of significant risk items (outlined below). However the financial investment required to proceed to next steps is significantly lower than is typically the case for NXP products (1-2 FTE engineers).

### How do we monetize it?

This is probably the largest risk associated with the project, as the techniques discussed below make use of sensor data, but are not implemented on the sensor themselves. How do we win, and KEEP, customers?

Option 1:

Publish an SDK and hope for the best. This would capture mindshare, but probably not a lot of revenue.

Option 2:

Create and license a reference design that includes:

embedded application including feature extraction, machine learning, tracking, anomaly detection and reporting functions

Windows or Cloud-based (Azure or IBM BlueMix) GUI for device configuration and monitoring

Option 3:

Manufacture and sell PCBs implementing the reference design above

The NTI will include investigation into alliance and partnerships that may be leveraged *for technical infrastructure* to successfully monetize NXP’s solution and determine the minimal and optimal offerings NXP can offer to participate in this market.

Any of the above could involve partnering with one or more companies as preferred customers.

### Technical Risks

* Optimal choice of statistical features vary from project to project. Our best option is to provide a fairly large selection of features, coupled with a logical ORing of univariant anomaly detectors based upon each of those features. The downside of this approach is that it cannot take advantage of n-dimensional patterns in the data.
* Real systems exhibit features which drift in time. Differentiating between normal and abnormal drift is an area that is likely to become one of “those nagging problems”

### Support Risks

Options (2) and (3) in 2.1.1 above would likely have a higher support burden than earlier projects such as sensor fusion. Machine learning is still new to most engineers over age 30. We’ll have to provide training and set expectations with regard to performance.

## Learnings to date

We now have a good understanding of the following topics:

* Fault types which can occur in rotating machinery
* FFT
* Wavelets
* Statistical Feature Extraction
* Supervised Machine Learning Techniques
  + Clustering: Gaussian Mixture Models, K-nearest neighbors
  + Support Vector Machine
  + Decision Trees
* Tools:
  + Matlab (programmatic and Classification Learner)
  + Microsoft Azure
  + IBM SPSS and BlueMix
  + Knime

Key findings have been:

* It is very difficult to get ‘failing data” for supervised learning. This has caused us to redirect our efforts to single class anomaly detection – i.e., Learn normal operation and flag any departures from that.
* Normal machine drift over time implies our models must be adaptive – which makes them more complex and less sensitive.
* Selection of features to be used for machine learning is still somewhat of an art form. This makes “canning a solution” more problematic.

The last two bullets above are the major technical challenges (excluding manufacturing) remaining for us.

# Target customers & volume; target markets

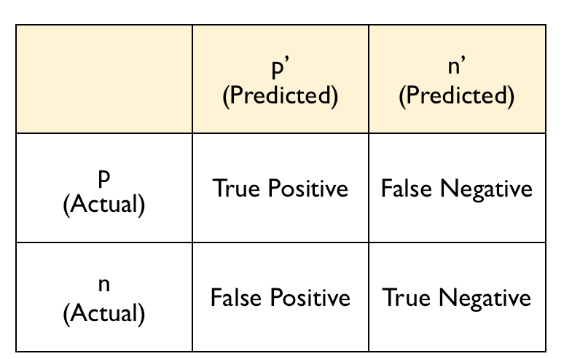
Light industrial and white goods markets.

# Target timeline

Prototypes: 3Q2017.

# Quality Criteria

Machine learning problems are graded via a “decision matrix”:



Ideally, all test cases lay on the diagonal. However machine learning is a statistical process, and there will almost always be off-diagonal components. Recurring false positives result in a “boy who cried wolf” scenario. False negatives result is missed opportunities to service equipment before it becomes a critical issue.

# Performance requirements

Solution should fit and execute on an M4F-class MCU.

See above.

# Project Board required roles

<???>

# Version History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Modified by | Description of Changes |
| 0.1 | 8/16/2016 | Michael Stanley | Initial version for internal review |
| 0.2 | 8/17/2016 | Ian Chen |  |
| 0.3 | 9/1/2016 | Michael Stanley | Compromise version between 0.1 and 0.2. |
| 0.4 | 9/1/2016 | Michael Stanley | Compromise version between 0.1 and 0.2. |