KCF technologies

# CODE CHALLENGE

KCF Analytics Software Engineering

336 S Fraser Street, State College, PA 16801 | 814.867.4097 | www.kcftech.com

# KCF Analytics Software Engineering Code Challenge

The goal of this challenge is to assess the applicant's ability to solve a real-world problem using physical reasoning, statistics, and computer programming. You will produce:

- source code implementing a solution to this challenge.
- a written technical brief supplying the details requested in this challenge.

## Background

KCF vibration transducer packages encapsulate two vibration sensors that measure acceleration in units of *g* (9.8 m/s^2). At regular intervals, they wake up, acquire 4096 measurements at a rate of 8192 Hz, transmit their results, and return to idle. (Depending on sensor settings, the sample rate or length may be different, but most are configured to take 4096 measurements at 8192 Hz.) We refer to each of these periodic sets of measurements as a *burst*. Each transducer package records vibration on two axes: X, parallel to the base of the transducer package, and Y, perpendicular to the base.

KCF places one or more of these transducer packages on a machine. The location of a transducer package is called a *monitoring point*. The acceleration data received from the sensors at each monitoring point can tell us a lot about the health of the machine; we and our customers use this information to make maintenance decisions.

## The Challenge

Develop a program that can determine whether a machine is on or off, depending on data from KCF vibration sensors.

This program takes a zip archive as its input and produces a two-column CSV file as its output. The first column will be timestamped, and the second will be a number, where 0 represents "off" and 1 represents "on". The output file should not have a header row. The output file should be ordered by timestamp.

A zip archive of sample data has been prepared for this challenge. Your program will be evaluated by running it on another file with different data in the same format.

## The Data

This challenge is provided with a zip archive containing data from several KCF vibration sensors. These sensors are on different monitoring points on the same machine.

The files in the zip archive have the following organizational convention:

```
7bfbb0a6-f81e-40a5-8aca-2d9b00cda47a/Motor Outboard/C3F0BA61/Y/0621.csv
               |                              |             |    |   |
       unique asset id           monitoring point name  |  axis  |
                                                        |   month and year
                                                 sensor serial number
```

There are eight files in the zip archive, all from June 2021, all from the same machine.

Internally, one of the CSV files looks like this:

```
1622525136160,8192,4096,0.0676,0.0091,-0.139, ...
       |        |    |    |
       |        |    |    sample data ...
    timestamp   | sample length
            sample rate
```

Each row of the CSV file corresponds to one burst of sample data:

- timestamp: measured in milliseconds since midnight, January 1, 1970, UTC (the Unix epoch).
- sample rate: measured in Hz.
- sample length: number of measurements in the burst.
- sample data: the remaining values in each row are the acceleration measurements.

An important physical characteristic of the accelerometers used by KCF is that they report the force of gravity as experienced by the sensor in the direction of measurement, as well as acceleration due to mechanical vibration. Depending on the orientation of the sensor, this shows up in the data as a constant offset of between -1*g* and +1*g*.

Each CSV in the archive contains some bursts that were acquired when the machine was off and some bursts that were acquired when the machine was running.

## The Program

Your program must be written in Python 3.11 and must pass pylint>=3.0 and mypy>=1.7 with no errors or warnings. Please provide source code (you may retain the license to this code, so long as we are allowed to run it on our test data for evaluation.) In addition, please explain exactly how to run your program:

- operating system (Windows or Linux, or either)
- library requirements
- program invocation instructions

We have access to Microsoft Visual Studio. We also can install open-source software on Windows or Linux. We will be unable to evaluate a program that requires the purchase of any additional licenses.

We have provided a skeleton of a program for you to modify. You are free to alter or completely replace the provided program skeleton as needed to suit your requirements.

## The Technical Brief

When your program is completed, please prepare a technical brief. This document should have 1-3 pages of text, and it may additionally contain up to 10 supplemental figures.

The technical brief should address the following questions:

- What is the relationship between the data in the zip archive and the data in the on/off column? Include formulas as appropriate.
- Why does this work? Provide tables or supplemental figures in support of your answer. In addition, support your answer with physical reasoning.
- What assumptions must be satisfied for this approach to work on other machines?
- From the data, what can you infer about the production schedule at the plant where this machine is installed?

## The Evaluation

We will run your program on data from a similar machine and evaluate whether the output makes sense.

We will read your source code and evaluate it for maintainability and performance (in that order.)

We will read your technical brief and evaluate it on clarity of presentation and the strength of your reasoning. You don't have to be right, but we want to see you make a plausible case.

## Thank You

We appreciate your effort! This is a challenging programming task that involves real-world data and a real-world problem. We hope that completing this task gives you a sense of what it's like to work in the Analytics department at KCF and that you enjoy the experience!