

## A1.05-9348 - Data Mining for IVHM Using Ensembles of Decision Trees

PI: John Trenkle Michigan Aerospace Corporation (MAC) - Ann Arbor, MI

### Identification and Significance of Innovation:

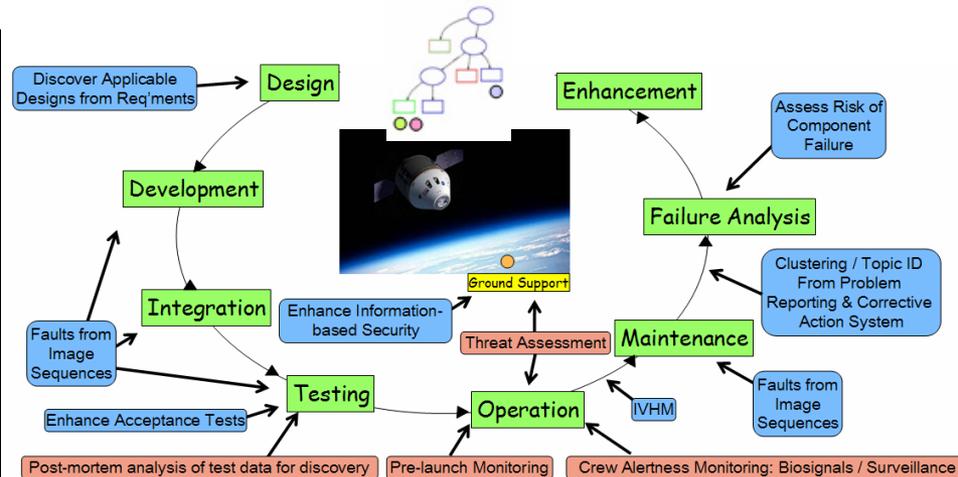
- o MAC's SPADE (Sparse Processing Applied to Data Exploitation) addresses a significant portion of the topic's goals and has functionality that can address many applications within the aircraft life cycle.
- o Highly quantizes input data and uses MAC's novel techniques for constructing Ensembles of Decision Trees (EDTs) to develop extremely accurate diagnostic and prognostic models for classification, regression, clustering, anomaly detection and semi-supervised learning.
- o EDTs have capability to explain *why* a given decision is made.
- o Significant advantages to the SPADE approach: 1) completely data driven; 2) training and evaluation are faster than conventional methods; 3) operates effectively on huge datasets (> billion samples X > million features), 4) proven to be as accurate as state-of-the-art techniques in many significant real-world applications.

MAC will work closely with NASA domain experts to determine how the SPADE approach may be applied to particular areas in the vehicle life-cycle.

Expected TRL Range at the end of Phase I Contract (1-9): 3-4

### Technical Objectives and Work Plan:

- Objective 1:** Demonstrate the depth and breadth of current SPADE functionality on actual data
  - Objective 2:** Work with NASA and corporate partners to delineate appropriate applications in the vehicle lifecycle
  - Objective 3:** Delineate the roadmap for enhancing SPADE to address the specific needs of various applications
- Task 1:** Demonstrate Current State of SPADE for Real Applications
    - o Work with internal data or data of opportunity to demonstrate capabilities
  - Task 2:** Requirements Analysis for Potential Applications
    - o Characterize storage, memory and computational aspects of current system
    - o Devise deployment plans for candidate platforms (multi-core desktops boxes, airborne platforms, FPGAs, GPUs...)
  - Task 3:** Roadmap of Benefits to Enhance Potential Applications
    - o with specific apps in mind describe algorithmic enhancements



Candidate areas for applying MAC's ensemble-based data mining technology

### NASA and Non-NASA Applications:

**NASA:** MAC's SPADE data mining system has a large potential market in both government and civil aviation as well as for other arenas with costly and complex vehicles such as marine craft. The need for next-generation data mining tools for aid in lifecycle issues for aircraft/spacecraft/satellites/ships is now widely recognized by both the private and public sectors, as exemplified by the scope of the solicitation for this program.

**Commercial:** data fusion approaches to computer network security, intelligence, and public health monitoring; real-time quality control and damage detection for continuous physical processes; text stream monitoring for news, email, IMs; financial event detection; sales opportunity/threat identification; micro-climate change monitoring using digital imagery; gene expression profiling for medical diagnosis and understanding of diseases; proteomic data analysis and pattern recognition for medical diagnosis and biomarker discovery...

Firm Contacts: SpaceX, GMV Space Systems, Boeing