NRC·CNRC

Institute for Information Technology

Applications of data mining in modern maintenance

IAGT/NRC collaborative forum on *Challenges and Opportunities in Future Gas Turbine*

October 20th, 2008

Sylvain Létourneau

NRC·CNRC
Institute for Information Technology

Knowledge from data at the NRC's IIT

Help organizations maximize the benefits of the data that they collect

- perform scientific research in: machine learning, data mining, statistics, artificial intelligence, soft computing, information retrieval, information systems, and natural language processing
- preferred application areas: Equipment Health Management, Bioinformatics, Health, Web Mining
- today's talk limited to Equipment Health Management

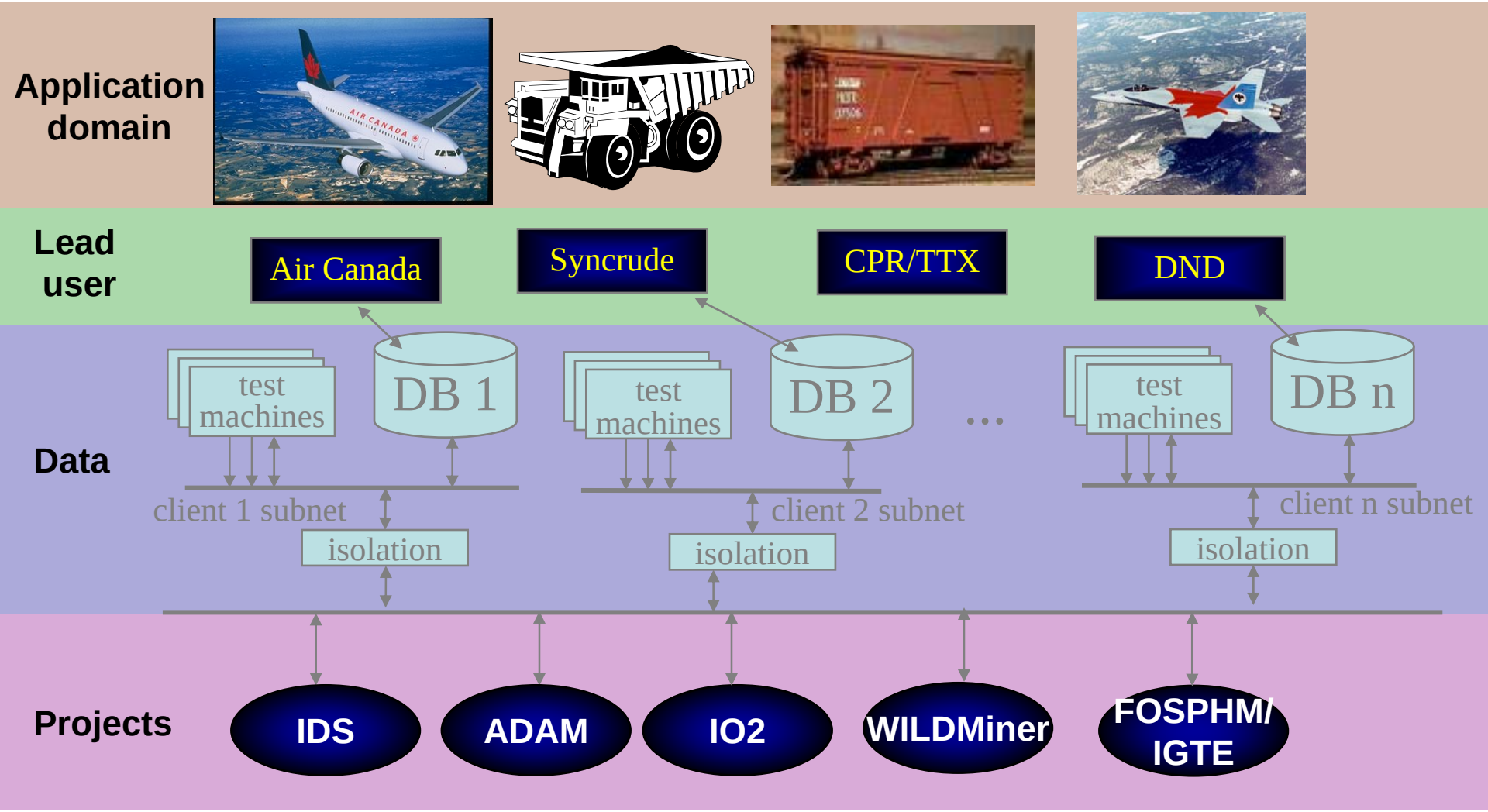
NRC·CNRC
Institute for Information Technology

Areas of contribution

- Data management
  - cleansing of data
  - storing data into data bases and data warehouses
  - providing secure and convenient access to the data
- Data-driven modeling, reasoning, and simulation
  - summarizing raw data into manageable information through statistics
  - extracting useful insights through data mining, statistics, machine learning,…
  - using learned models to enhance simulation and background knowledge
  - fusing information and automate/support decision making
- Computing
  - developing high performance computing platforms and tools to perform required computations
- Software tools
  - developing tools that integrate and streamline the data analysis process
  - developing software to put results from data analysis at work

NRC·CNRC
Institute for Information Technology

Data warehouses for the various projects at NRC-IIT



NRC·CNRC
Institute for Information Technology

Integrated Diagnostic System (IDS) - overview

- Timeline
  - 1991- thorough study of commercial aircraft maintenance and assessment of potential benefits of IT technologies to support decision making
  - 1993- official start of the IDS project with partners such as Canadian Airlines, Air Canada, GE, Lockheed Martin, CMC, …
  - 1996- prototype installed at Air Canada for 3 month trial (lasted 3 years…) - 69 aircraft (A320/319)
  - 2000- commercialization
- Scope and technologies
  - uses of Artificial Intelligence (AI) techniques for diagnosis and recommendation of repair actions
  - focuses on Air Canada's line technicians for A319/A320

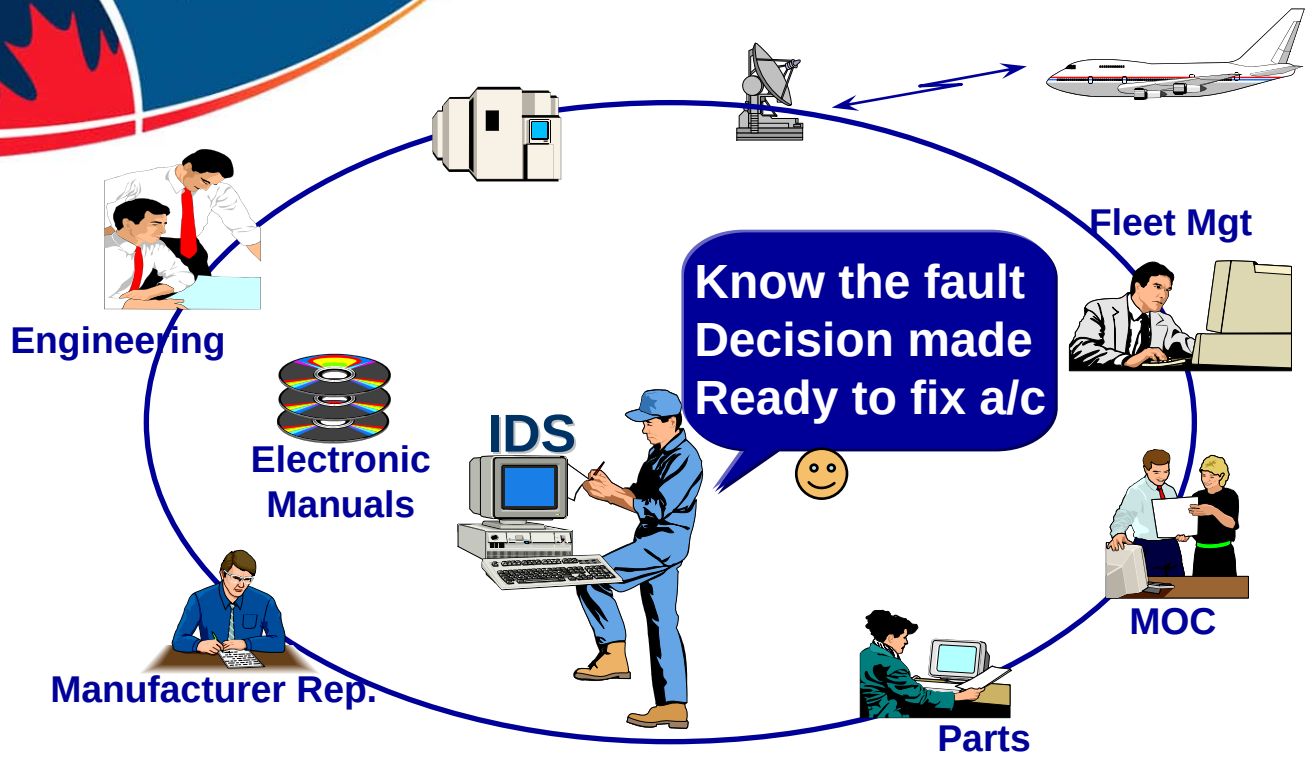
Line technician's world

35 minutes

- aircraft at gate
- passenger offload
- snag recognition
- consult MEL
- consult TSM
- consult aircraft history
- carry out test (BITE)
- fault isolation (TSM)
- parts required?
- order parts
- rectification (AMM)
- certification

NRC·CNRC
Institute for Information Technology

IDS - concept



- While the aircraft is still in the air, IDS:
  - clusters relevant WRN, FLR, and snag (pilot) messages
  - identifies probable causes based on TSM and provide links to corresponding TSM pages
  - assesses MEL conditions (GO, NOGO, GOIF, …)
  - displays relevant maintenance history
  - suggests fixes based on similar cases
- facilitates communication between staff from various departments

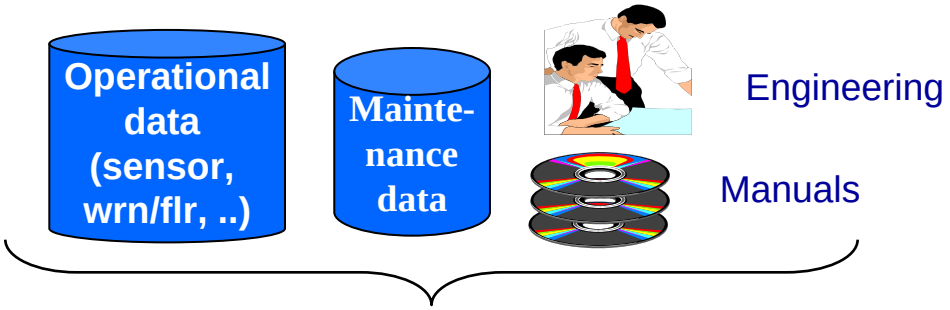
NRC·CNRC
Institute for Information Technology

IDS - technologies

- Case-Based reasoning (CBR)
  - to identify FLR and WRN messages. Case-retrieval includes advanced string matching to compensate for data errors
  - to suggest repairs based on similar past experiences
- Rule-based reasoning
  - to represent knowledge contained in TSM manual. Development of tools to automatically generate these rules
  - to cluster FLR and WRN messages and identify of probable faults based-on TSM
  - to assess MEL condition
  - to aggregate information related with temporal proximity or textual similarity
- Appealing and efficient end user application interface

NRC·CNRC
Institute for Information Technology

Data mining for equipment health management



Build models to detect and/or predict failures

Failure Models

Decision support system

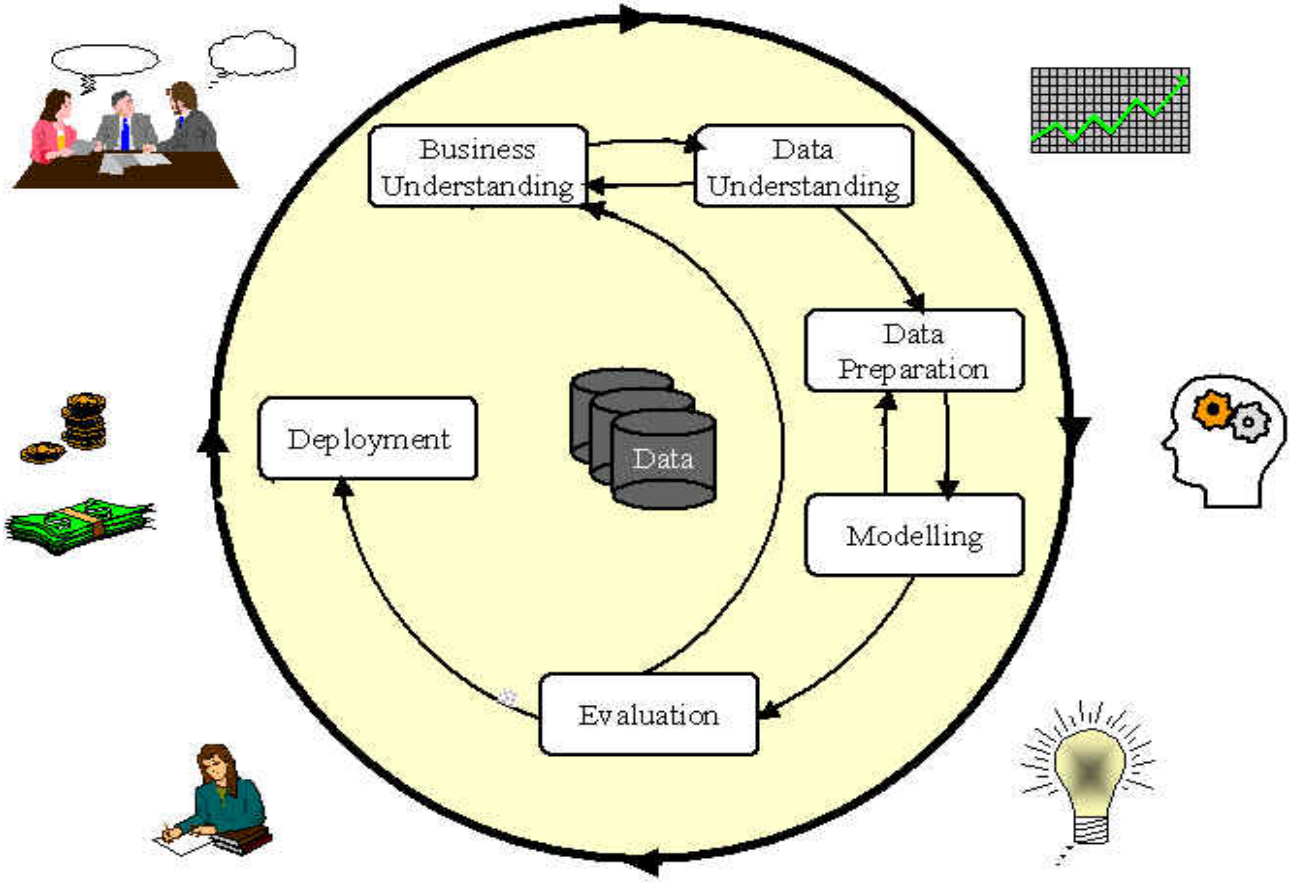
Eng #1, Fin 217
Very high fuel-flow

APU, Fin 204
Starter likely to fail within 15 days

- Use sensor data, maintenance data, and domain information
- Two kinds of models:
  - component failure predictions
  - abnormal behavior detections

Data mining overview: definition and process

"Data mining (DM) is the process of discovering useful and previously unknown knowledge from historical or on-line data"



NRC·CNRC
Institute for Information Technology

Data mining overview: modeling techniques

- Classification/prediction/trend analysis

  classical statistics (discriminant analysis, time series analysis, etc.), decision and regression trees, (naïve) bayes, probabilistic networks (Bayesian networks/markov networks), artificial neural networks, fuzzy-logic, rule induction, k-nearest neighbor/case based reasoning, inductive logic programming, rough sets, genetic algorithms, evolutionary systems, …

- Association

  association rules, inductive logic programming, ...

- Clustering

  hierarchical and probabilistic cluster analysis, fuzzy cluster analysis, conceptual clustering, kohonen feature maps, ...

- Plus many hybrid methods

- Selection of techniques is done on a case by case basis according to types of modeling task and characteristics of the data

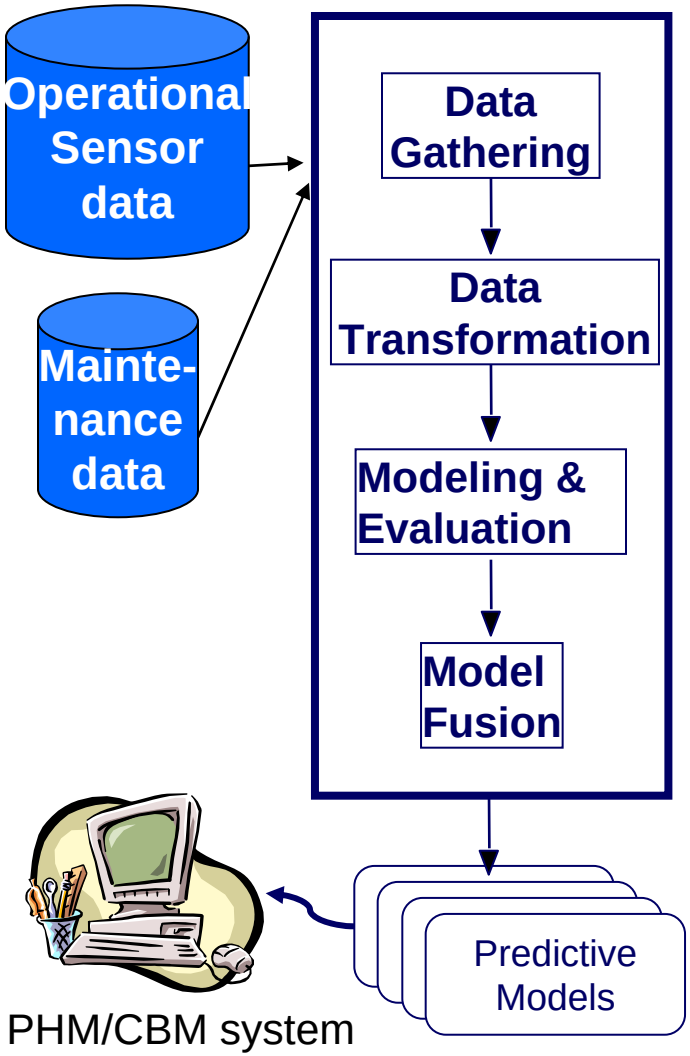
NRC·CNRC
Institute for Information Technology

Data mining for health management: challenges

- Integration of time information
  - most data mining tools assume no order in observations
- Selection of relevant data
  - Several datasets and not all data from a given dataset is relevant for a given problem
- Processing for noise and contextual information
- Labeling of the data
  - Class parameter typically not given, it needs to be determined
- Evaluation of the models
  - Typical evaluation process and functions are not adequate
- Fusion of models
  - Often needs to integrate several models to achieve desired results

NRC·CNRC
Institute for Information Technology

Data mining methodology for health management

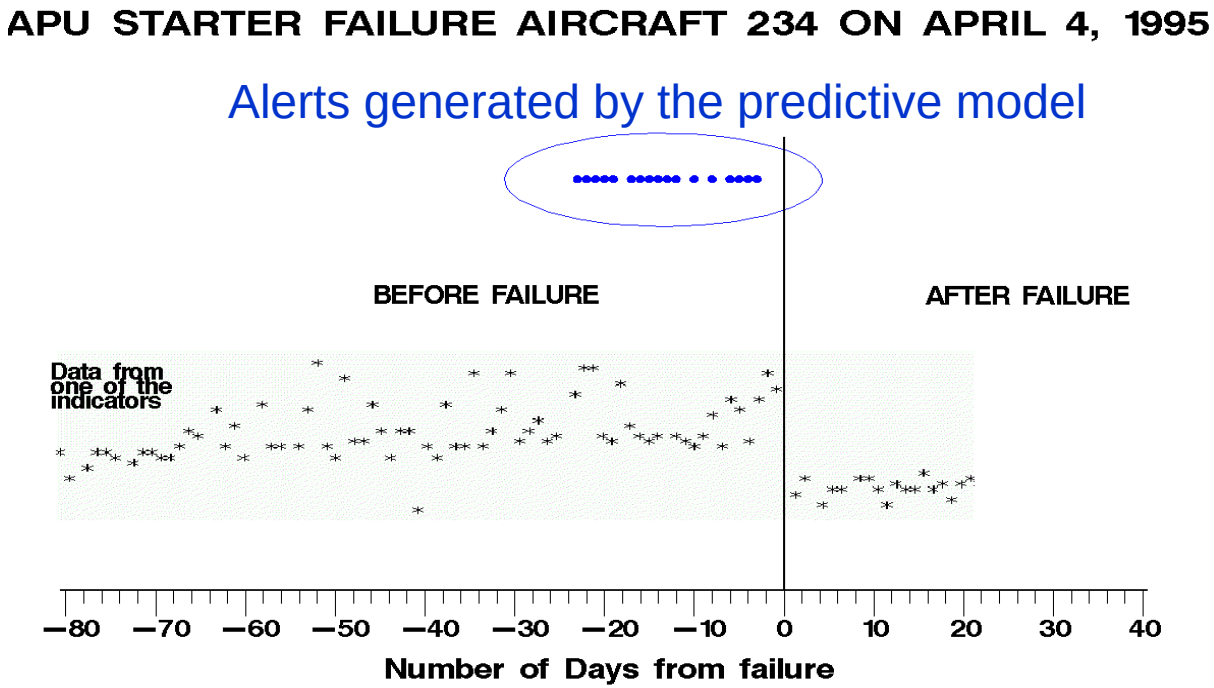


Iterative process, main tasks:

- retrieving past failures information
- data selection
- data labeling
- feature extraction
- building data mining models
- leave-one batch out evaluation methodology
- application specific scoring function
- heterogeneous model stacking
- data mining learning of meta models

Example 1: predicting APU starter failures

- **Objective** predict A320 APU starter failures between 1 and 30 days in advance to avoid delays
- Illustration with a particular case

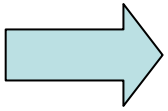


  - the model should generate alerts in the target period before the failure

Example 1: predicting APU starter failures (2)

- **Data**
  - 6 years of maintenance and APU sensor data; 5 years for building the models and the remaining year for evaluation (training/model building)
  - 69 occurrences of starter replacements over first 5 years (64 from A320 and 5 from A319) and 17 in the remaining year (testing/model evaluation)

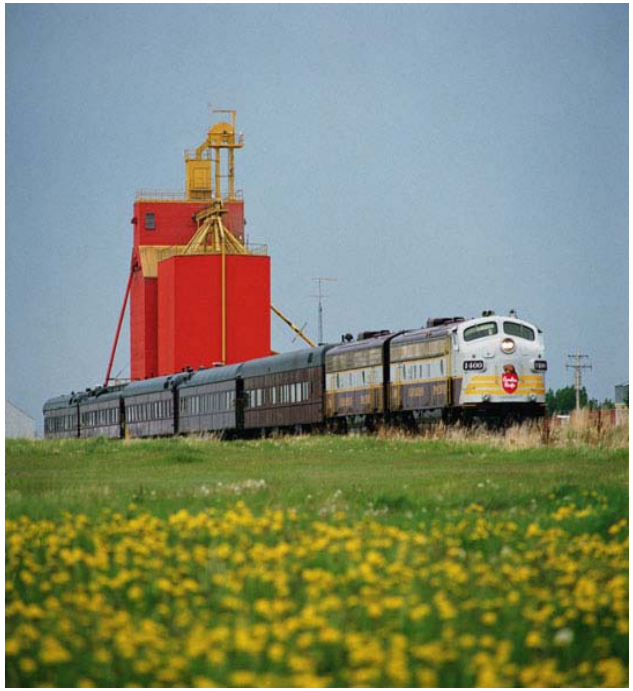
- **Results** (on testing data – failures during the last year of the study)
  - A320: models generated alerts for 11 occurrences out of 12 (total of 113 alerts, 5 potential false alerts)
  - A319: models generated alerts for 2 occurrences out of 5 (total of 61 alerts, 27 good, and 34 to be validated)



These models could be used to avoid many delays at the gate

NRC·CNRC
Institute for Information Technology

Example 2: predicting train wheel failures

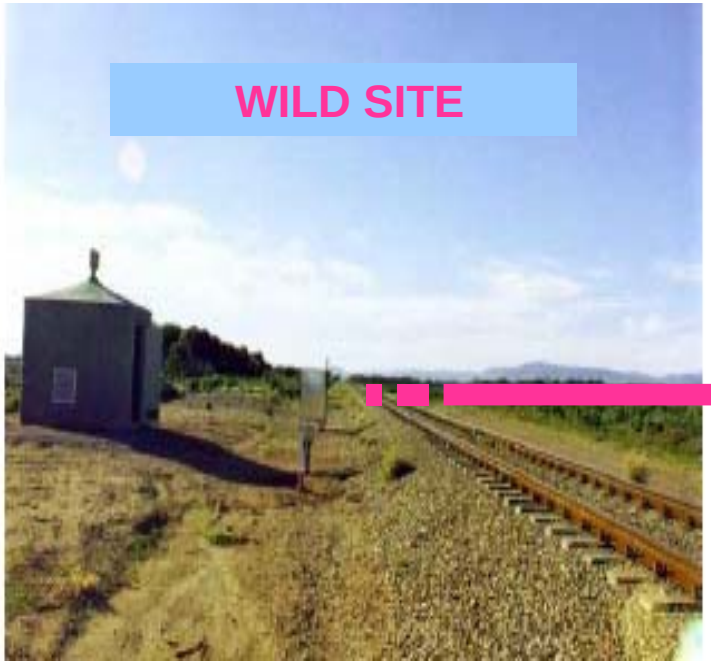


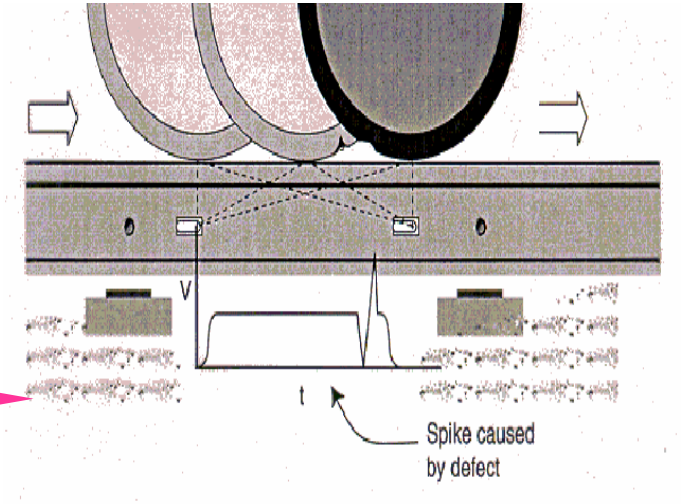
- Train wheel failures:
  - account for more than 50% of train derailments
  - cause significant disruptions of railway operation: force changes in schedule, reduced throughput, cause delays
  - increase maintenance cost ($65M /per year for wheel repairs)
  - reduce life of rail (5000 broken rails per year)
  - are more and more frequent due to increased load and speed

- **Objective** use IIT's data mining technology to predict wheel failures and avoid disruptions during operation
- **Data** maintenance data plus WILD (Wheel Impact Load Detector) data

Example 2: predicting train wheel failures (2)

WILD technology





Wheel Impact Load Detector

- Output for each wheel: dynamic load, nominal weight, speed, direction
- 22 WILD sites located at strategic location on the Canadian railway network
- Current policy is to stop a train when impact > 140 kips

Example 2: predicting train wheel failures (3)

- **Result**

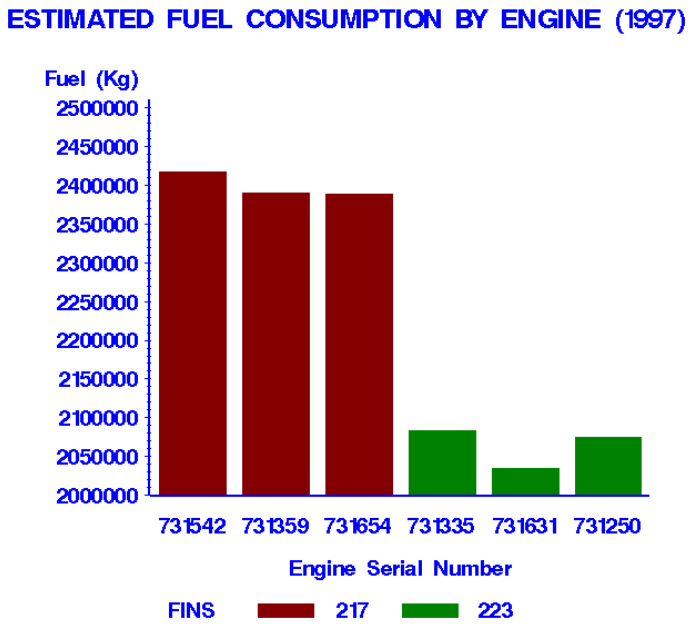
Meta-Model Name	*Version of Algorithms*	*Model Score*	*False Positive Rate*	*Problem Detection Rate*
m_1^c	Decision trees	698.5	0.08	0.97
m_2^c	Decision trees with costMatrix	650.9	0.08	0.97
m_3^c	Naïve Bayes with costMatrix	643.4	0.12	0.98
m_4^c	Naïve Bayes	622.7	0.13	0.98

- good coverage and reasonable rate of false positives
- by comparison, threshold-based approaches
  - often fail to provide timely alerts
  - generate a much higher rate of false positives for a given detection rate
- simple methods are not feasible due to variability observed between wheel failure cases

NRC·CNRC
Institute for Information Technology

Example 3: detection of abnormal behavior

- **Objective** detect abnormal behaviors in the operation of aircraft engines by closely monitoring key parameters for trends or sudden shift in performance
- **Data**
  - 5 years of Engine Cruise Reports from a Air Canada's fleet of A320
  - monitoring of EGT, Fuel Flow, N1 vib, and N2 vib
- **Result**
  - AC 217 consumes too much fuel whatever engine you put on it
  - over 1 yr, AC 217 consumes about $207000 more in fuel than AC 223 (assuming 5hr crusing per day and $0.31/kg of fuel)
  - In 2000, Airbus diagnosed an airframe problem on that aircraft and performed major repairs…



NRC·CNRC
Institute for Information Technology

Hybrid modeling: knowledge and data driven

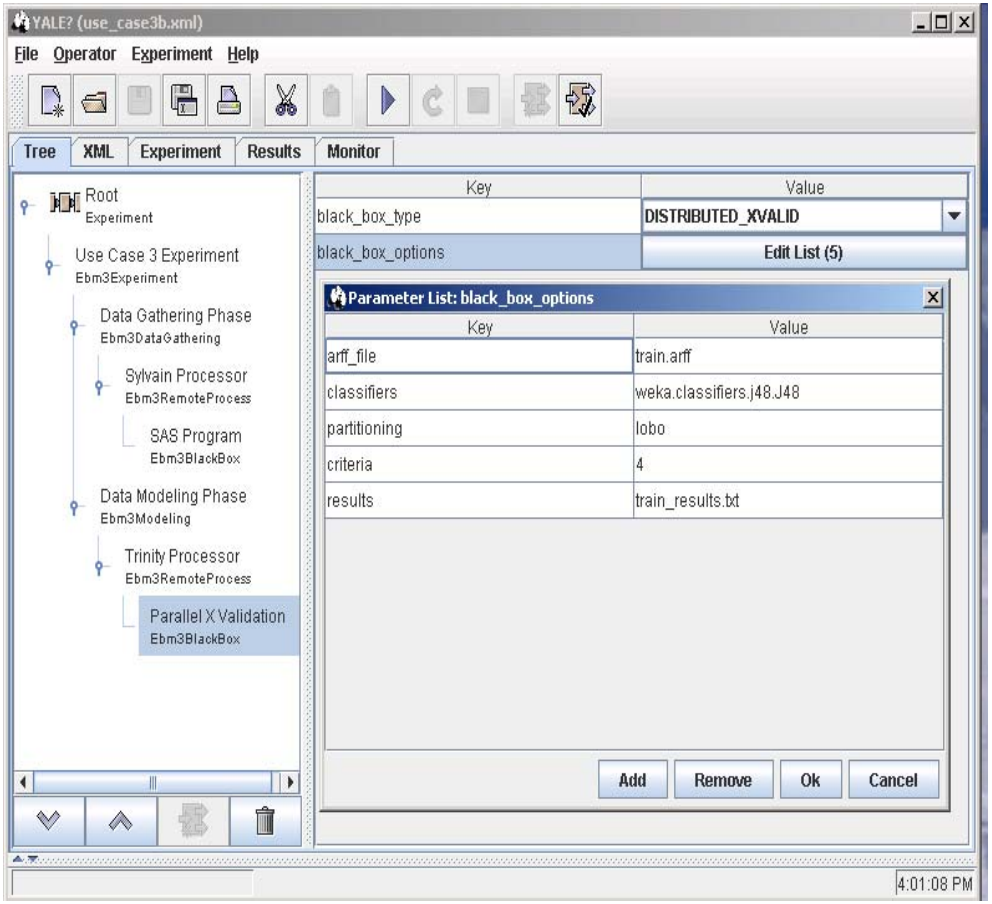
- *Data-driven* approaches construct generalized models that capture the relationships between the input and output data of a given process
- *Knowledge-driven* or *physics-based* approaches construct models that try to explain the physics underlying a given process
- Why do we need both?
  - to help deal with lack of data or noise in the data
  - to help deal with lack of knowledge or build initial models for complex systems
- Examples of integration of knowledge in data-driven modeling:
  - train wheel failure predictions
    - normalization of load measurements based on speed and nominal weight
    - selection of sensor readings based-on expected propagation of impact due to wheel failures
  - aircraft engine health assessment
    - normalization of measurements based on generator load
    - understanding operating envelope under various conditions (compressor maps)

NRC·CNRC
Institute for Information Technology

Software tools for model development and deployment

NRC·CNRC
Institute for Information Technology

EBM3 system: environment for building models



Functionalities:

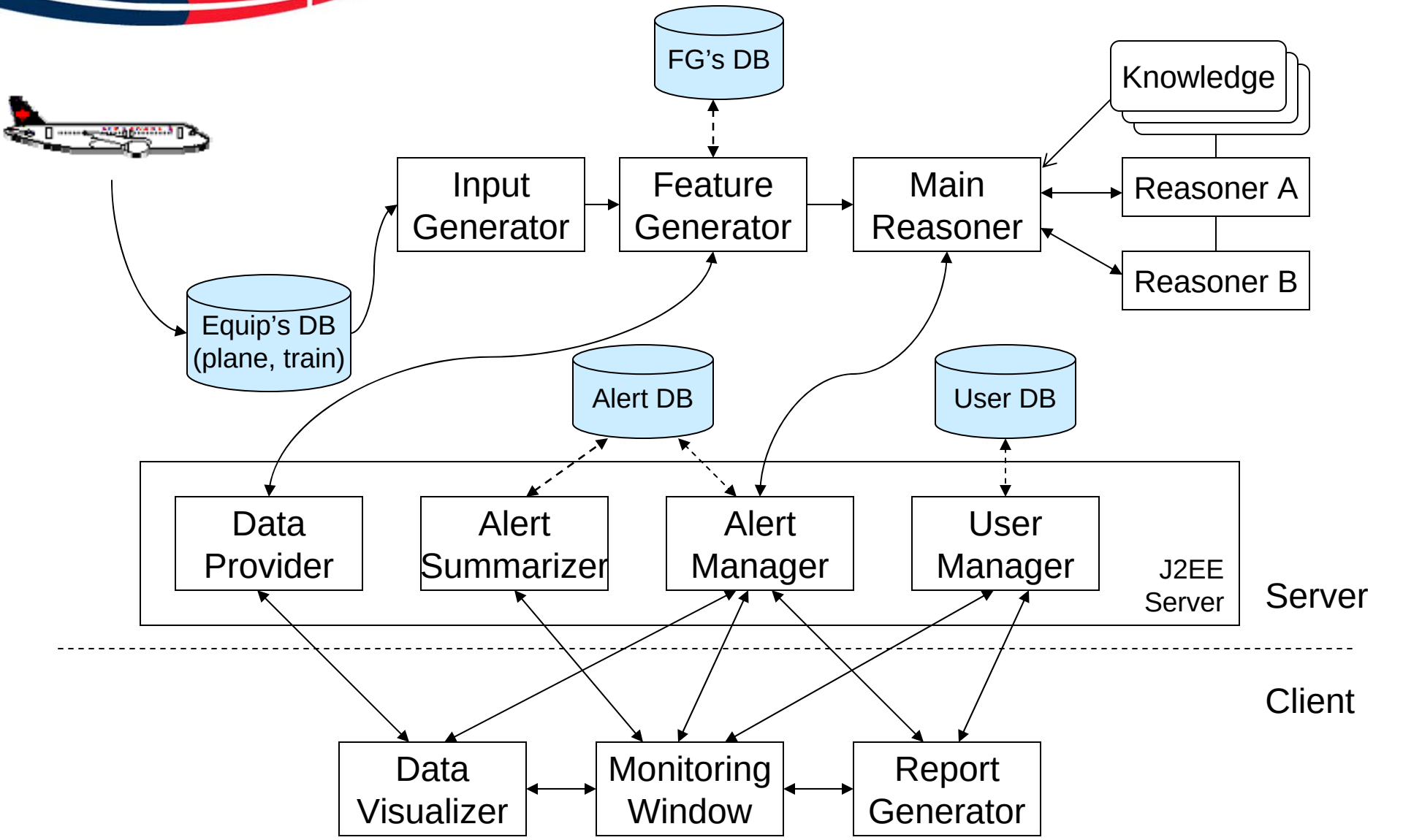
- streamline model building, deployment, and documentation process
- support for variety of application development environments
  - SAS, Java, R, Matlab, Perl…
- support for different computing environments
  - Windows, Linux, Parallel and Distributed Computing
- support integration of new techniques in the model building process
- support cooperative work

Experiments performed so far:

- time-to-failure estimation
- cost curve analysis
- export models to EDM3

NRC·CNRC
Institute for Information Technology

EDM3 system: researching and demonstrating results



NRC·CNRC
Institute for Information Technology

Conclusion and practical considerations

- information technologies can help optimize maintenance (e.g., increase availability) by facilitating the integration and processing of huge amounts of data, knowledge, and expertise
- technology development/implementation requires:
  - an adequate business case and resources (data)
  - strong participation from end-users (operators)
  - adequate data
  - a multi-disciplinary and iterative approach
  - an open software environment that complies with industry standards to allow integration of systems, technologies, and data

NRC·CNRC
Institute for Information Technology

Questions

1. How to foster the participation of operators in the development of health management technologies?

2. What are the barriers to data sharing and what could be done collectively to augment the quantity and quality of data available for technology development?