THE VALUE OF DATA SCIENCE STANDARDS IN MANUFACTURING ANALYTICS

SOUNDAR SRINIVASAN
BOSCH DATA MINING SOLUTIONS AND SERVICES

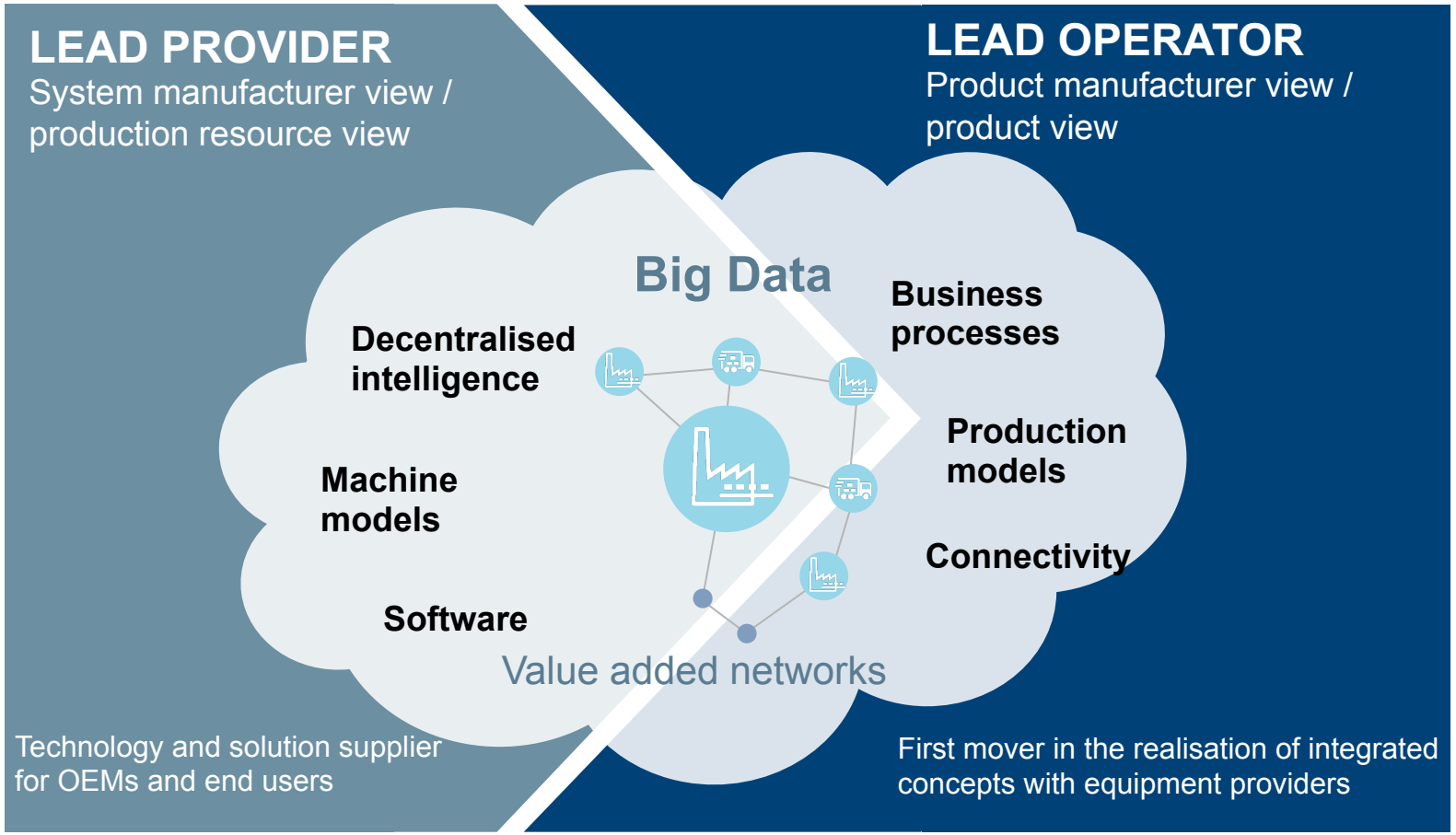
BOSCH

Data science standards in manufacturing analytics
Outline

- Bosch's dual role in advanced manufacturing/Industry 4.0
- The need for standards in predictive analytics
- Case study in the use of PMML at Bosch
- How to improve existing standards?

Data science standards in manufacturing analytics

Two perspectives for Bosch on Industry 4.0



BOSCH

Data science standards in manufacturing analytics
The need for standards in predictive analytics

- **Bosch's interest in Standards for Manufacturing Analytics**
  - As an user/operator
    - Vendor independence
      - » Interoperability and Standardization of data collection, storage, retrieval, and presentation
      - » Data-driven verification and validation for improving efficiency and quicker scaling
      - » Use of best practices and standards to improve quality and traceability
      - » Model auditing and update
  - As a provider
    - Interoperability and Standardization
    - Sharing of success stories and best practices
    - Drive adoption of data-driven modeling, V&V
      - » Bosch is a leading participant in ASME's initiative on **verification and validation** for advanced manufacturing
    - Creation of neutral testbeds and certification agencies

Data science standards in manufacturing analytics

Join the manufacturing analytics community

- **Predictive Modeling in Manufacturing Analytics Challenge**
  - Kaggle Competition to be launched on August 17th, 2016
  - Focus on improving product quality as a binary classification problem (0.6% in one class)
    - 1 year of a product manufactured in large volumes and probably in your car
    - Complete assembly and testing data
    - 3 million samples, 4000 features,
  - Public testbed for manufacturing data science innovation
- **IEEE Big Data for Advanced Manufacturing Special Symposium**

  2016 IEEE International Conference on Big Data

  Dec 5 – Dec 8 2016 @Washington D.C., USA

  http://cci.drexel.edu/bigdata/bigdata2016/SpecialSymposium.html

  August 31, 2016: Results due for the manufacturing data challenge

  Sept 20, 2016: Due date for full symposium papers submission

Data science standards in manufacturing analytics

Analytics success stories in manufacturing

Test and Calibration Time Reduction

- Prediction of test results
- Prediction of calibration parameters

Scrap Costs Reduction

- Early prediction from process parameters
- Descriptive analytics for root-cause analysis

Warranty Cost Reduction

Prediction of field failures from

- Test and process data
- Cross-value stream analysis

Yield Improvement

- Benchmark analysis across lines and plants
- Pin-point possible root causes for performance bottlenecks (OEE, cycle time)

Predictive Maintenance

- Identify top failure causes
- Predict component failures to avoid unscheduled machine down-times

Data science standards in manufacturing analytics

Case Study: Test Time Reduction

Business Objective:

Reduce test and calibration time in the production of mobile hydraulic pumps

Impact

***35% reduction** in test and calibration time via accurate prediction of calibration and test results*

Data science standards in manufacturing analytics

Case Study: Test Time Reduction

Layout of the assembly line

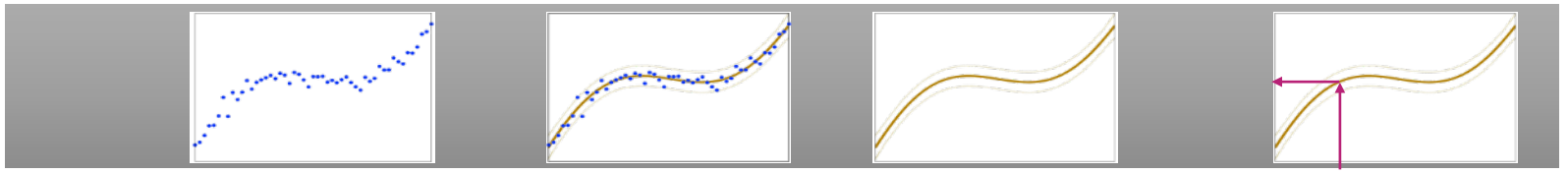
Problem:
Bottleneck Test Benches

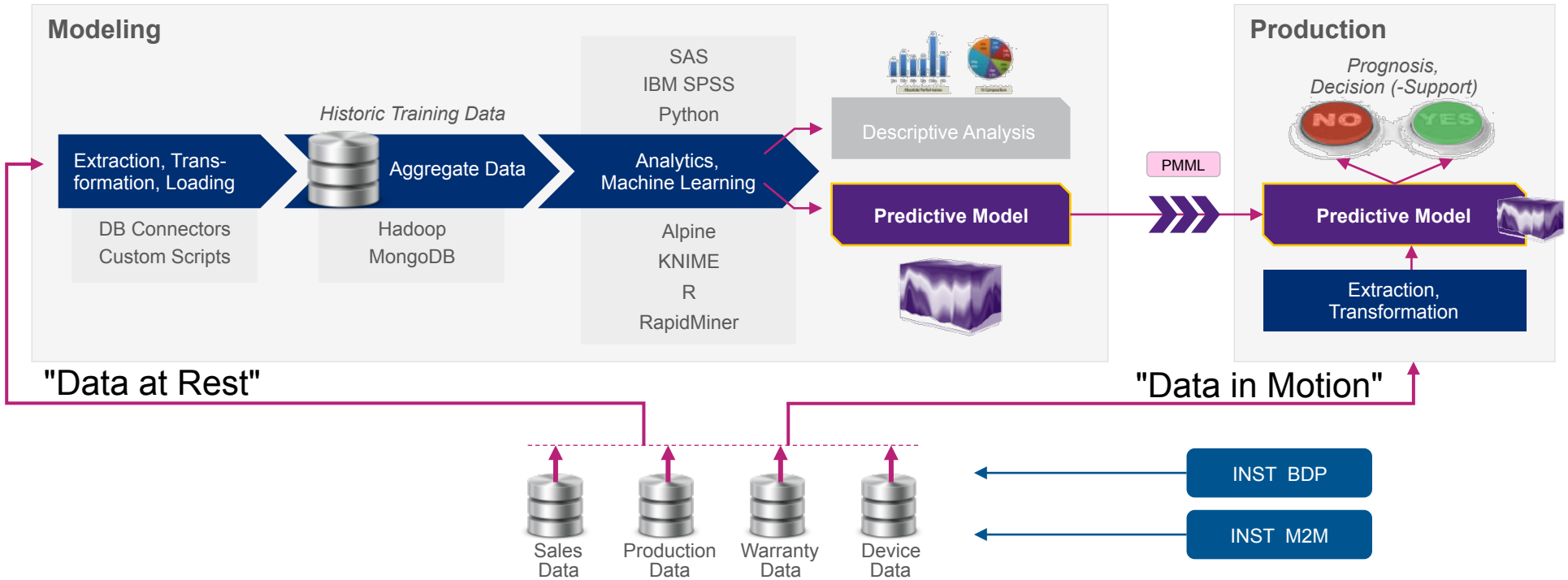
Approach:

1) Identify candidate tests for removal
2) Identify test '**groups**' run in parallel
3) Use **feature selection** methods to identify **least important** test measurements.
4) Remove least important test measurements (saving test time)
5) Train a predictive model to predict test outcome from remaining measurements.

Data science standards in manufacturing analytics

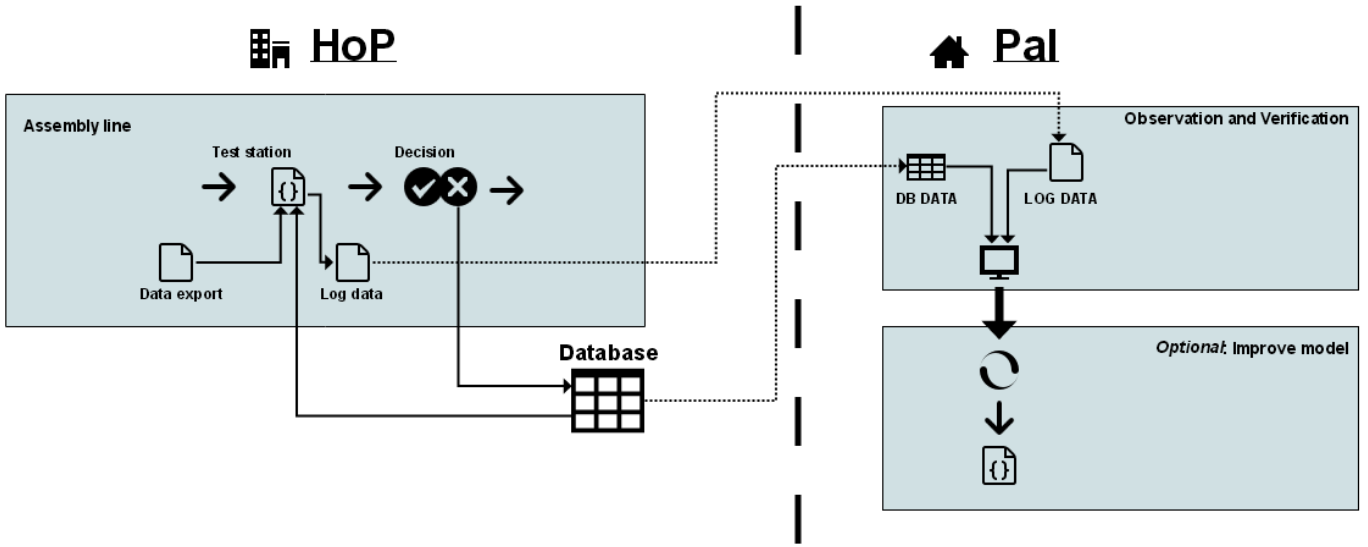
Our analytics information workflow





Data science standards in manufacturing analytics
Deployment using PMML

- Model (Boosted Trees) developed in R
- Implementation time ~1 month
  - Proposed a client-server architecture using the PMML implementation by ADAPA
  - No installation required at the client



BOSCH

Data science standards in manufacturing analytics
Alternatives to PMML use

- Deployment using R-server
  - Not robust enough for continuous and low latency deployment
  - Additional memory overhead for low cost machines in manufacturing
  - Need to create scoring logs
- End-to-end deployment using other freeware or commercial analytics software
  - Local installation required
  - Need to recreate solutions
  - Learning overhead for data scientists
  - Licensing costs

Data science standards in manufacturing analytics
Summary of first impressions in using PMML

- Vendor independence
- Freedom of development tools for the data scientist

- Each vendor implements PMML differently
- Model coverage is limited
  - Adapa had to be extended in our application; many thanks to Zementis for a quick response
- Commercial solutions have better support, but come at a higher cost

Data science standards in manufacturing analytics
How to improve existing standards?

- Certification of compliance by DMG
- Keep up with the innovation in modeling paradigms
- Standards have to cover the complete analytical workflow
  - ETL
  - Model creation
  - Model deployment
  - Validation
  - Interpretation and uncertainty quantification
  - Versioning and traceability
- Consideration of development and deployment environments