Data Science Services by Omron

How to get the full value from your factory floor data

CONTENT

WHITEPAPER

Industry 4.0 (i4.0) and Industrial Internet of Things (IIoT) are buzz words for several years and i4.0 concepts are actually implemented on more and more machines. A huge amount of data becomes available: machine data, data of the production process and data regarding the manufactured product. Big Data has entered the factory floor.

Data is nowadays easily collected and stored, but in most cases the 'data pipeline' stops here and there is hardly any value extracted from the data. The 'data pipeline' is often not completed in a proper way so that the right person(s) can easily exploit the value inside the data. It is a challenge to extract the value from the huge stream of data and not to drown in the flood. Only collecting and storing of data is not enough to monetize the investments in the i4.0 and IIoT infrastructure.

Getting the maximum value out of the data and keeping an overview of data streams nowadays goes beyond standard statistical methods and tooling. Manual analysis and creation of dashboards and reports is not sufficient. The dashboards become too complicated and are not showing the right information at the right time, in the right way, to be able to see at a glance what is going on and to be able to act. The routines implemented in a normal machine controller to observe the production process and to detect errors, are able to detect present deviations and problems, but are not suitable to predict future problems. Machine controllers are not suitable to combine all available information and to perform advanced analytics on it.

This is the reason that the discipline of (Industrial) Data Science emerged. Data Scientists are able to cope with the 3V's of big data (Fig. 1):
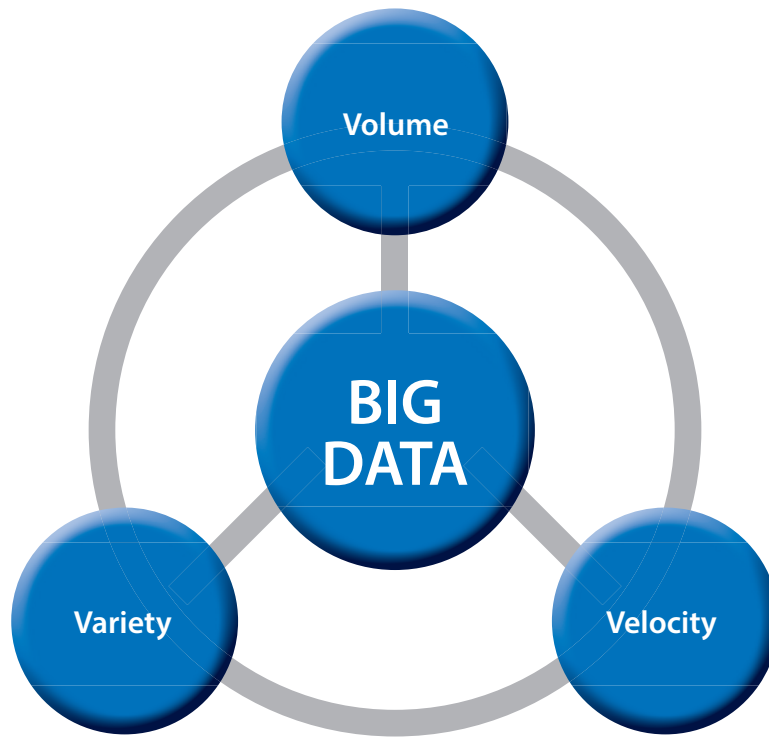
Fig. 1 – The 3 V's of big data

Volume: This is the characteristic of big data that most people think of first. A modern packaging machine can easily generate GigaBytes of data per day that is stored for an extensive period. For inspection machines, this can go up to TeraBytes per day. Storage of these amounts of data is not a problem, but utilizing it is a different story.

Variety: A machine nowadays is not only producing a few numbers of data, but the nature of the data is much broader: not only the measurement is stored, but also the raw values of sensors and other 'meta data' of the sensors and actuators. Not only the inspection results are stored, but also the taken pictures. Data can also be generated by the operator of the machine, like tact times and feedback in the form of text or even speech.

Velocity: This term not only refers to the velocity of the data is generated, raw data from sensors is typically read every millisecond and needs to be treated as streaming data, but it also refers to the expected speed of analyzing the data. Updating a dashboard once per day or once per hour is not adequate. An operator wants to be notified of potential problems right away to prevent product loss and downtime. Ideally, the machine is notified in real-time, so it can correct itself automatically within the same product cycle.

Sometimes a fourth V is mentioned in the definition of Big Data:

Veracity: Data can be faulty due to a problem in the sensor or other device, data can be missing or data can be recorded in an unconventional or old-fashioned way. This can highly affect further analysis and lead to incorrect conclusions if the veracity is not specifically taken into account.

Data Science Project Approach

Industrial Data Science is a very new discipline and there is no one-size-fits-all solution (yet). Each solution and application need tailored data analysis and modelling to obtain the maximum result.

Data Scientists at Omron follow a standard approach (Fig. 2) to obtain the best project results and to manage the expectations. The approach is based on the CRISP-DM model. CRISP-DM stands for Cross-Industry Standard Process for Data Mining and is widely used.

Fig. 2 – Project phases of the standard approach used by Omron Data Scientists.

Phase 1: Preparation

The problem or request is refined with all stakeholders and domain experts to come to a well-defined project goal.

The machine and/or production process is analyzed on a high level to get an overview of what data is already available and what data collection needs to be implemented to obtain the project goal. A first set of data is collected and analyzed during this data understanding phase as a kind of feasibility study.
At the end of the preparation phase, a business proposal is written to give insight in the expected generated value and realistic ROI.

The preparation phase is the most important phase. A Data Science project will never succeed without having the goal very clear. Data analysis always lead to interesting discoveries and there is a risk that the project will drift away because of this.

Phase 2: Analysis and Application Development

The data is collected over a longer period to get a representative reflection in the data of the machine and process behavior. A data pipeline (Fig. 3) is developed that depending the project goal contains the following stages:

- Data Collection: Data is collected from various sources, ranging from raw sensor data to information from MES systems.
- Data Preprocessing: The collected data is prepared for the analytics step by transforming, merging and cleaning of the data.
- Data Analytics: The developed analytics algorithms and trained machine learning models are applied.
- Apply Results: The results and conclusions from the data analytics stage are made available by for instance a visualization tailored to the audience and situation or by sending feedback to the machine.
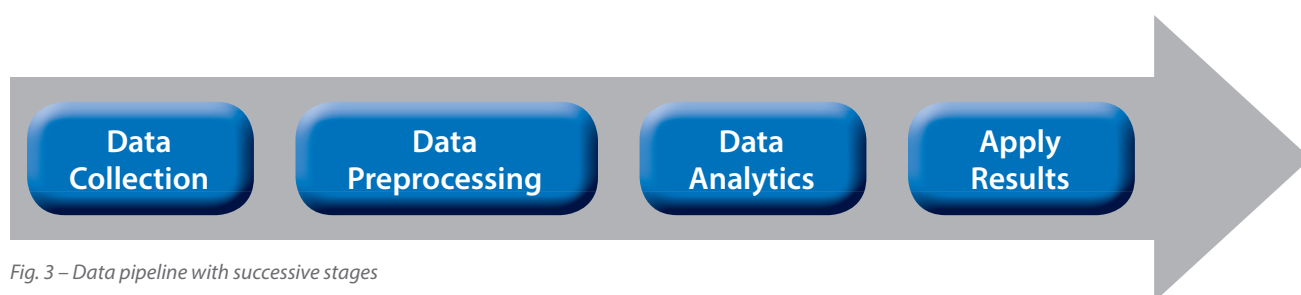

Fig. 3 – Data pipeline with successive stages

The necessary machine learning models are trained and validated together with the rest of the data processing steps. Once the validation is successful, an application is developed containing the data pipeline that can be easily deployed and executed.

Phase 3: Evaluation

The application is deployed in the production environment and the performance and business results are assessed. The previous project phases are revisited if the performance is not as expected.

Phase 4: Maintenance

Production processes change and machine behavior changes (due to updates or wear) over time. A regular revalidation of the solution is necessary to ensure that the solution stays connected with the reality and keeps its value. The amount of available data grows over time and often better models can be developed with more data. This can also be a valid reason to periodically revisit the existing (machine learning) models.

Application Example

A data-driven solution does not always have to involve the use of fancy machine learning models or advanced artificial intelligence. Processing the data the right way and providing the right information, at the right time, in the right way can sometimes already do the trick. An example of such a Data Science project performed at the Omron Manufacturing of the Netherlands (OMN) factory is explained in this paper.

The project was performed at the Surface-Mount Technology (SMT) lines (Fig. 4), where electronic components are mounted and soldered onto printed circuit boards (PCB's). This process consists of five steps.
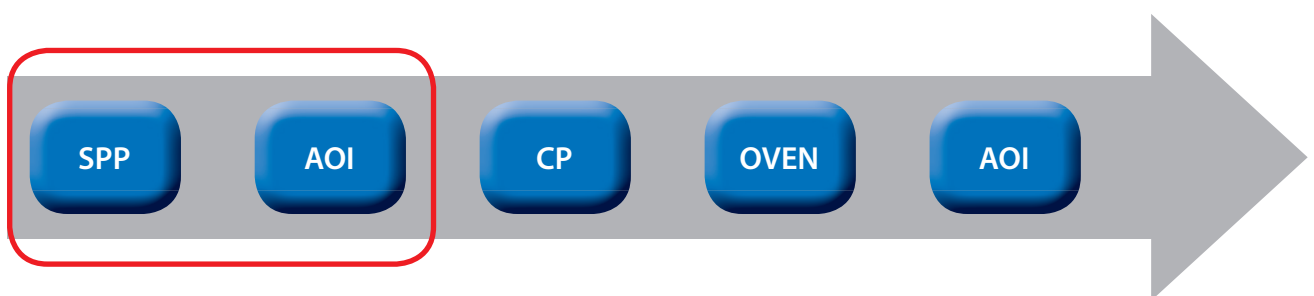

Fig. 4 – The Surface-Mount Technology line. This paper focusses on the first two process steps.

The first step is the Stencil Printing Process (SPP) where solder paste is applied onto the PCB on the locations where the electronic components will have electrical contact with the PCB. The second step is Automatic Optical Inspection (AOI). Here the Stencil Printing Process is evaluated, and the PCB is checked for (potential) quality problems. Next the electronic components are placed on the PCB before the PCB enters the oven, where the solder paste melts, and the actual soldering takes place. The PCB runs through a second AOI machine at the end of the line for a last visual quality check.

It is estimated that the root cause for 50 – 70% of the quality issues can be traced back to the Stencil Printing Process, so this process step is closely monitored by the line operators. The AOI machine gives a clear signal to the operator by a light pole and buzzer when it detects a potential quality issue. It takes an operator a lot of time to analyze an alarm from the AOI and to judge if an alarm is a real problem. The AOI machine detects all deviations but not all deviations result into a quality issue at the end of the line.

The SMT team contacted the Data Science team with the request to come up with a data-driven solution that reduces the time it takes an operator or even the process specialist to analyze solder application problems.

It appeared that the Solder Printing machine logs a lot of data that was not used at all at that time. First idea was to start utilizing this data for the analysis of solder printing problems. A data pipeline was developed that collects all log files from the Solder Printing machine, preprocesses the data and stores it in a convenient format in a database in the local cloud. Several machine learning models were trained on this data. A neural network

was able to predict the Solder Printing quality with an accuracy of 95%. This is a very good performance for a machine learning model in general but turned out not to be a good solution: An accuracy of 95% is not accurate enough for the Solder Printing process. The model gave a lot of detailed insight to the SMT specialist of the general Solder Printing Process but did not help the operators to quickly resolve issues.

Secondly the focus was put on presenting the right data in an easy to understand manner to the operator. An operator may spend a lot of time finding the faulty area on the PCB, while detailed information of every soldering pad is logged by the AOI machine. This data was not used because it was not presented in a useful way to the operator. The operator also had no clear overview of the currently produced batch, only the results of the last PCB was shown.

A dashboard is developed that shows a heatmap of the complete PWB (Fig. 5). It shows a picture of the PWB with the locations with potential quality issues highlighted. The operator sees at a glance where he should focus his analysis on the board. The dashboard also shows the results of all PWB's of the batch. Areas where issues happen regularly are stronger highlighted on the heatmap. The operator draws the conclusion that the issues are not random, and action is taken on the Screen Printing machine, like cleaning the faulty area or adjusting the settings of the machine.

Fig. 5 – Operator dashboard showing solder application quality.

Now the operators can directly see where the problem is located, they need 60% less time to resolve issues.

The dashboard has 2 modes. The 'operator mode' only shows the heatmap of the currently produced batch. It is purely focused on supporting the operator with analyzing and resolving current quality issues. The 'specialist mode' enables the SMT specialist to drill down into the details and to compare different batches. He can analyze the occurrences over a longer period. The specialist also uses the dashboard to evaluate experiments to improve the production process by visually comparing the result of the experiment with regular production batches.

The data pipeline and dashboard are built mainly using open source tools: Python, Pandas, Plotly Dash and SQLite. This shows that open source is often sufficient, also for industrial data science applications. It also gives the opportunity for cheap experimentation: The data driven solution can first prove its value without having to invest a lot in a commercial tool set. If the solution has proven its value and needs to be expanded, the solution can be migrated to a commercial platform.

This solution also shows that a dashboard should be kept as simple as possible, only showing the key information needed by the target audience. Many visualization solutions display too much information, which often confuses people. The reason is that there are so many charts, tables, and many different colors and the audience gets lost and misses the most important message. For this reason, only the heatmap is shown to the operator and more detailed information is only available for the specialists.

Conclusion

It is challenging to use the potential of (big) data. Just collecting it and simply displaying some graphs is not sufficient as demonstrated in this paper. The valuable information needs to be extracted from the data and presented to the right audience, at the right time and in the right way.

The key is to put enough effort in the transformation process of the data into useful information. This has to be done in close collaboration between data scientists, who know how to tame the data and domain experts of the manufacturing process, who know the story behind the data. Only then a solution can be developed that not only looks interesting but is also used and brings value in the long run.

Would you like to exploit the full value of your industrial data or do you have a problem for which you believe the solution is hidden in your data? Contact Tim Foreman (**tim.foreman@omron.com**) or Dingeman Knaap (**dingeman.knaap@omron.com**)