NEWSIGNATURE

New Signature and
The Team Data Science Process

Technologies like Azure Machine Learning allow organisations to develop incredibly powerful predictive analytics capabilities, but the complexity of building and maintaining these models demand that we approach analytics with a thorough, rigorous approach.

The Team Data Science Process provides a tested framework for managing advanced analytics projects and delivering solutions that create sustainable value. Regardless of the discipline – from insurance risk forecasting to manufacturing process automation – data science must proceed through a formal process that allows experts to apply data and analytic technologies in a way that maximises the value of the results.

Lifecycle of the Data Science Method

Business Understanding

Understanding core business objectives and defining success is the first and most vital task in any data science engagement. Once success has been objectively defined, we can assess it in a valid way. This understanding informs all aspects of the project, dictating the data sources we access, the focus of analysis and data preparation, and the structure and content of models.

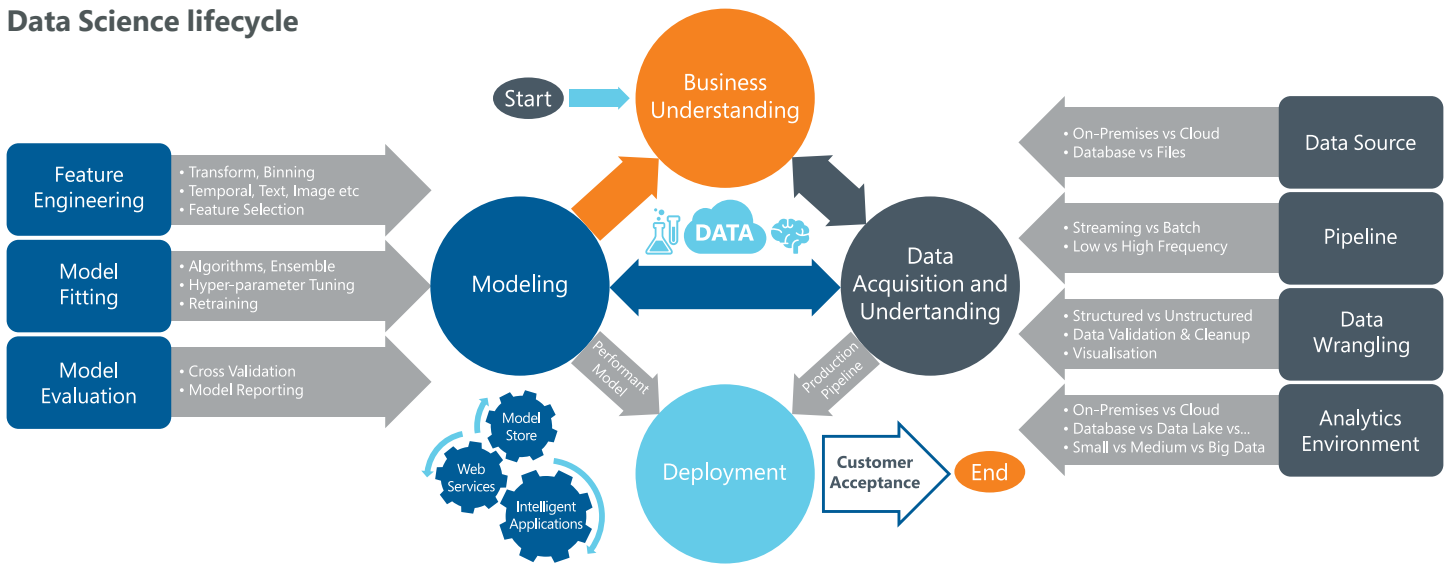## Related Technologies

Azure IoT Suite

Power BI

IoT Services

Process and Workflow Optimisation

Data Science lifecycle

Data Acquisition

Data acquisition involves assembling the data that will allow us to solve the problem we have defined. It involves a concerted strategy of identifying target data sources, understanding how they relate to our objectives, and performing an in-depth exploratory analysis. After identifying all relevant data, this data must be prepared and integrated into a dataset that supports predictive modeling. In addition to data collected and prepared for modeling, we must define and document any sources of real-time data that will be applied to a functioning predictive model for real-time analytics.

Modeling

Modeling is the process of applying machine learning techniques to a prepared dataset with the goal of producing a model that can predict something of value to the business. This is an iterative process that applies domain expertise, statistical techniques, and experimental procedures to develop models that are tuned to the specific objectives of the project. This involves generating valid predictors ("features") from the available data, selecting appropriate algorithms, training models, and rigorously evaluating results.

Deployment

Deployment allows us to deliver the value of predictive analytics to the business. Completed models can communicate with core business systems to make real-time predictions that improve outcomes, exposing predictive capabilities to the systems which can use them to improve efficiencies, deliver enhanced experiences, or expose valuable insights.

While the Team Data Science Process directly contributes to the success of Internet of Things (IoT) and Advanced Analytics projects, it also provides a mechanism for communicating insights. The team services approach and formalised workflows ensure that all documents and project artifacts that support the process are usable and understandable for other project teams.

The approach was instrumental to the success of a predictive analytics solution that New Signature developed for a Pennsylvania food manufacturer, which demonstrated a reduction in product variability and translated to significant and sustainable cost savings. In addition to the direct value that the solution created, the experience and application of a repeatable methodology helped to create a more connected enterprise that can access the value hidden within its data.