Building Azure Machine Learning Models

Eng Teong Cheah
MVP Visual Studio &
Development Technologies

Agenda

Team data science pocess lifecycle
Using regression algorithms
Using neural networks

Team data science process lifecycle

Team data science process

The Team Data Science Process(TDSP) is an agile, iterative data science methodology to deliver predictive analytics solutions and intelligent application efficiently.

TDSP helps improve team collaboration and learning.

It contains a distillation of the best practices and structures from Microsoft and others in the industry that facilitate the successful implementation of data science initiatives.

Team data science process

The goal is to help companies fully realize the benefits of their analytics program.

Data science lifecycle

This lifecycle has been designed for data science projects that ship as part of intelligent applications.

These applications deploy machine learning or artificial intelligence models for predictive analytics.

Exploratory data science projects or ad hoc analytics projects can also benefit from using this process.

Data science lifecycle

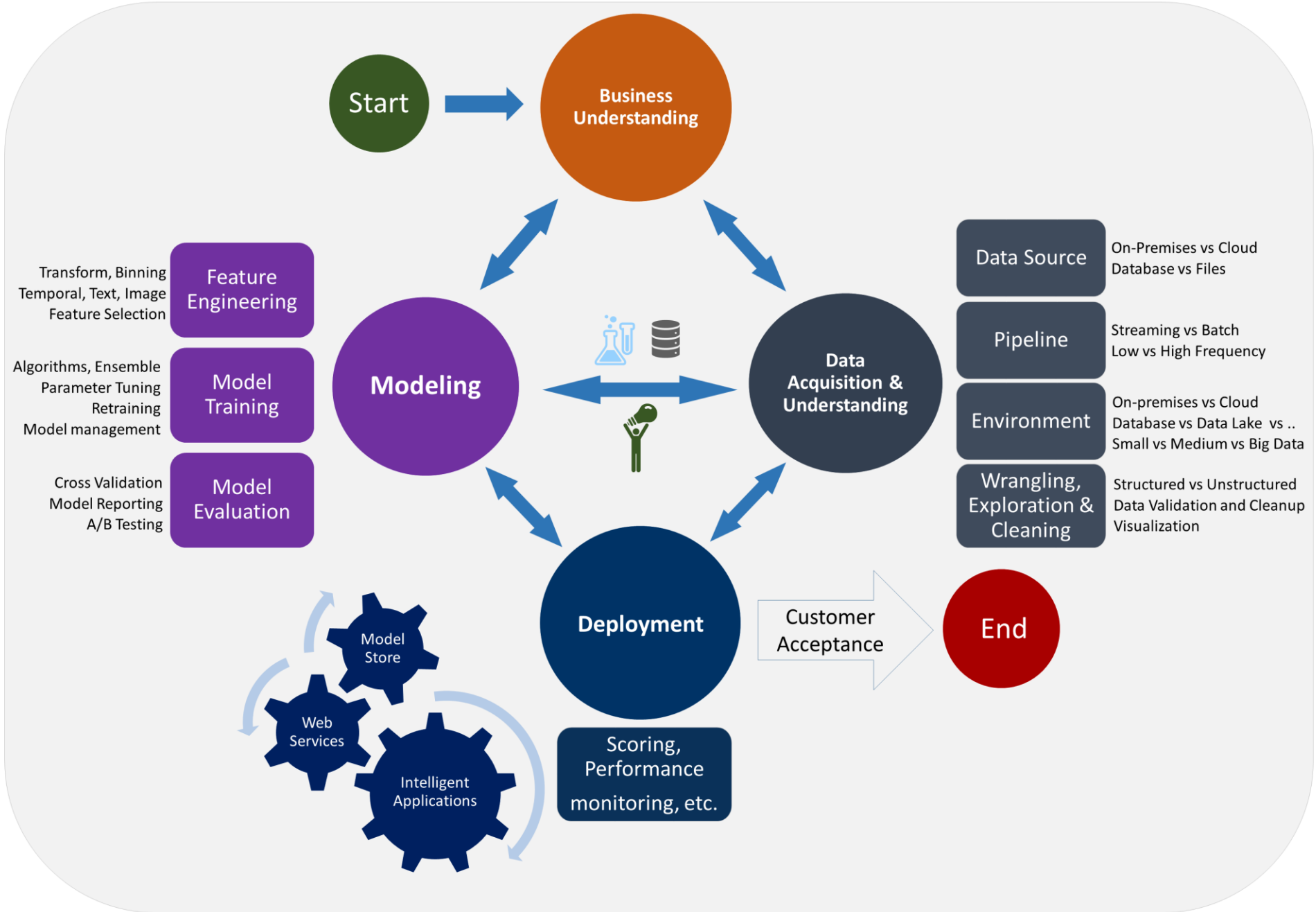
But in such cases some of the steps described may not be needed.

Data science lifecycle

The lifecycle outlines the major stages that projects typically execute, often iteratively:

- Business Understanding
- Data Acquisition and Understanding
- Modeling
- Deployment
- Customer Acceptance

Data Science Lifecycle



Using regression algorithms

Using regression algorithms

Regression is a methodology used widely in fields ranging from engineering to education.

For example, you might use regression to predict the value of a house based on regional data, or to create projections about future enrollment.

Regression task are supported in many tools: for example, Excel provides “What If” analysis, forecasting over time, and the Analysis ToolPak for traditional regression.

Using regression algorithms

The modules for regression in Machine Learning Studio each incorporate a different method, or algorithm, for regression.

In general, a regression algorithm tries to learn the value of a function for a particular instance of data.

You might predict someone's height by using a height function, or predict the probability of hospital admission based on medical test values.

Using regression algorithms

Regression algorithms can incorporate input from multiple features, by determining the contribution of each feature of the data to the regression function.

Ordinal regression

Ordinal regression is used when the label or target column contains numbers, but the numbers represent a ranking or order rather than a numeric measurement.

Predicting ordinal numbers requires a different algorithm than predicting the values of numbers on a continuous scale, because the numbers assigned to represent rank order do not have intrinsic.

Linear regression

Linear regression is a common statistical method, which has been adopted in machine learning and enhanced with many new methods for fitting the line and measuring error.

In the most basic sense, regression refers to prediction of a numeric target.

Linear regression

Linear regression is still a good choice when you want a very simple model for a basic predictive task.

Linear regression also tends to work well on high-dimensional, sparse data sets lacking complexity.

Bayesian linear regression

The Bayesian approach uses linear regression supplemented by additional information in the form of a prior probability distribution.

Prior information about the parameters is combined with a likelihood function to generate estimates for the parameters.

Bayesian linear regression

In contrast, the frequentist approach, represented by standard least-square linear regression, assumes that the data contains sufficient measurements to create a meaningful model.

Neural network regression

Although neural networks are widely known for use in deep learning and modeling complex problems such as image recognition, they are easily adapted to regress problems.

Any class of statistical models can be termed a neural network if they use adaptive weights and can approximate non-linear functions of their inputs.

Neural network regression

Thus neural network regression is suited to problems where a more traditional regression model cannot fit a solution.

Neural network regression is a supervised learning method, and therefore required a tagged dataset, which includes a label column.

Neural network regression

Because a regression model predicts a numerical value, the label column must be a numerical data type.

Boosted decision tree regression

Boosting is one of several classic methods for creating ensemble models, along with bagging, random forests, and so forth.

In Azure Machine Learning Studio, boosted decision trees use an efficient implementation of the MART gradient boosting algorithm.

Boosted decision tree regression

Gradient boosting is a machine learning technique for regression problems.

It builds each regression tree in a step-wise fashion, using a predefined loss function to measure the error in each step and correct for it in the next.

Thus the prediction model is actually an ensemble of weaker prediction models.

Poisson regression

Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts.

Fast forest quantile regression

Quantile regression is useful if you want to understand more about the distribution of the predicted value, rather than get a single mean prediction value.

Using neural networks

Neural network basics

A neural network structure consists of nodes that are organized in layers, and weighted connections (or edges) between the nodes.

The connections are directional, and each connection has a source node and a destination node.

Each trainable layer (a hidden or an output) has one or more **connection bundles**.

Neural network basics

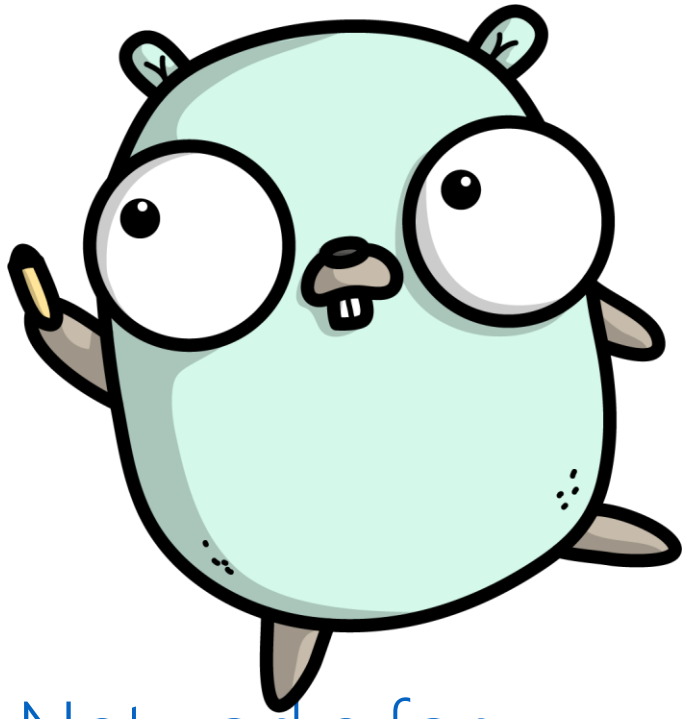
A connection bundle consists of a source layer and a specification of the connections from that source layer.

All the connections in a given bundle share the same source layer and the same destination layer.

Demo

Using neural networks in Machine Learning Studio

Resources



TutorialsPoint

Microsoft Docs

Lecture Collection | Convolutional Neural Networks for Visual Recognition(Spring 2017)

Python Numpy Tutorial

Image Credits: @ashleymcnamara

Thank you

Eng Teong Cheah
Microsoft MVP Visual Studio & Development Technologies
Twitter: @walkercet
Github: https://github.com/ceteongvanness
Blog: https://ceteongvanness.wordpress.com/
Youtube: http://bit.ly/etyoutubechannel