

Introduction to Data Mining/ Machine Learning



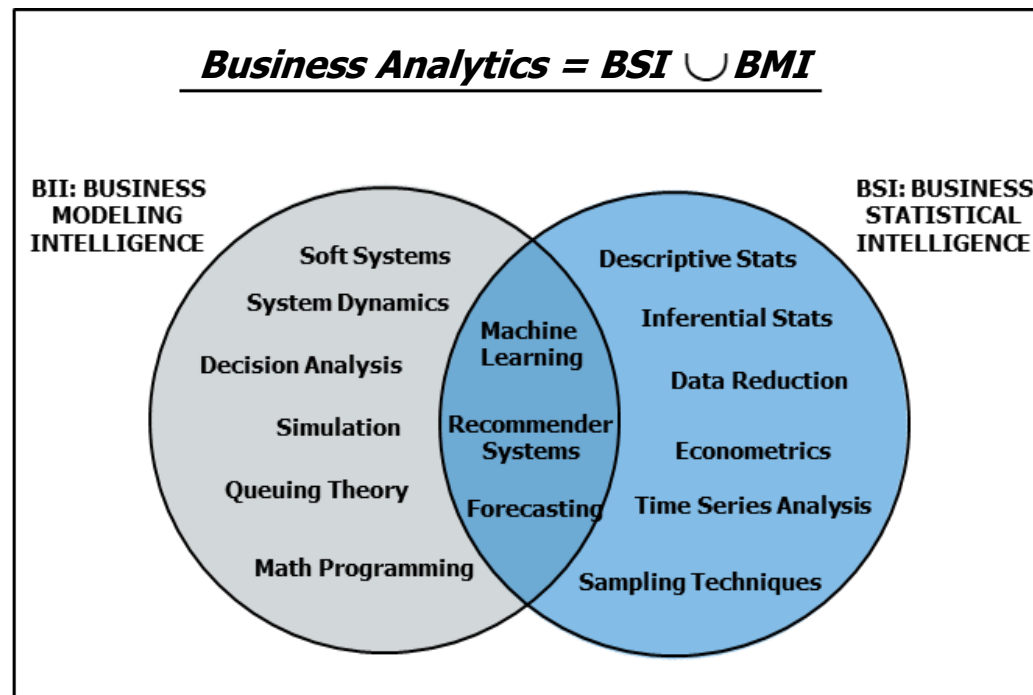
Providing software solutions since 1976

Andreas Zaras

Data Scientist

Business Analytics

*Business Analytics (BA), as defined by the International Institute of Analytics in 2010, is:
"The broad use of data and quantitative analysis to support the decision-making process...."*



Data Mining/ Machine Learning

Models or Algorithms

An algorithm is a recipe for using logic and mathematical operations to solve a problem.

A machine learning algorithm is often referred to as a learner.

Machine learning is a method of data analysis that automates analytical model building. It's considered a branch of artificial intelligence based on the idea that machines can learn and adapt through experience.

Data Mining/ Machine Learning

Considerations

Interpretability

The interpretability of a machine learning model is the degree it can be understood and explained by people.

Transparent, Simple, Informative. Trustworthy

Automated Machine Learning (AML)

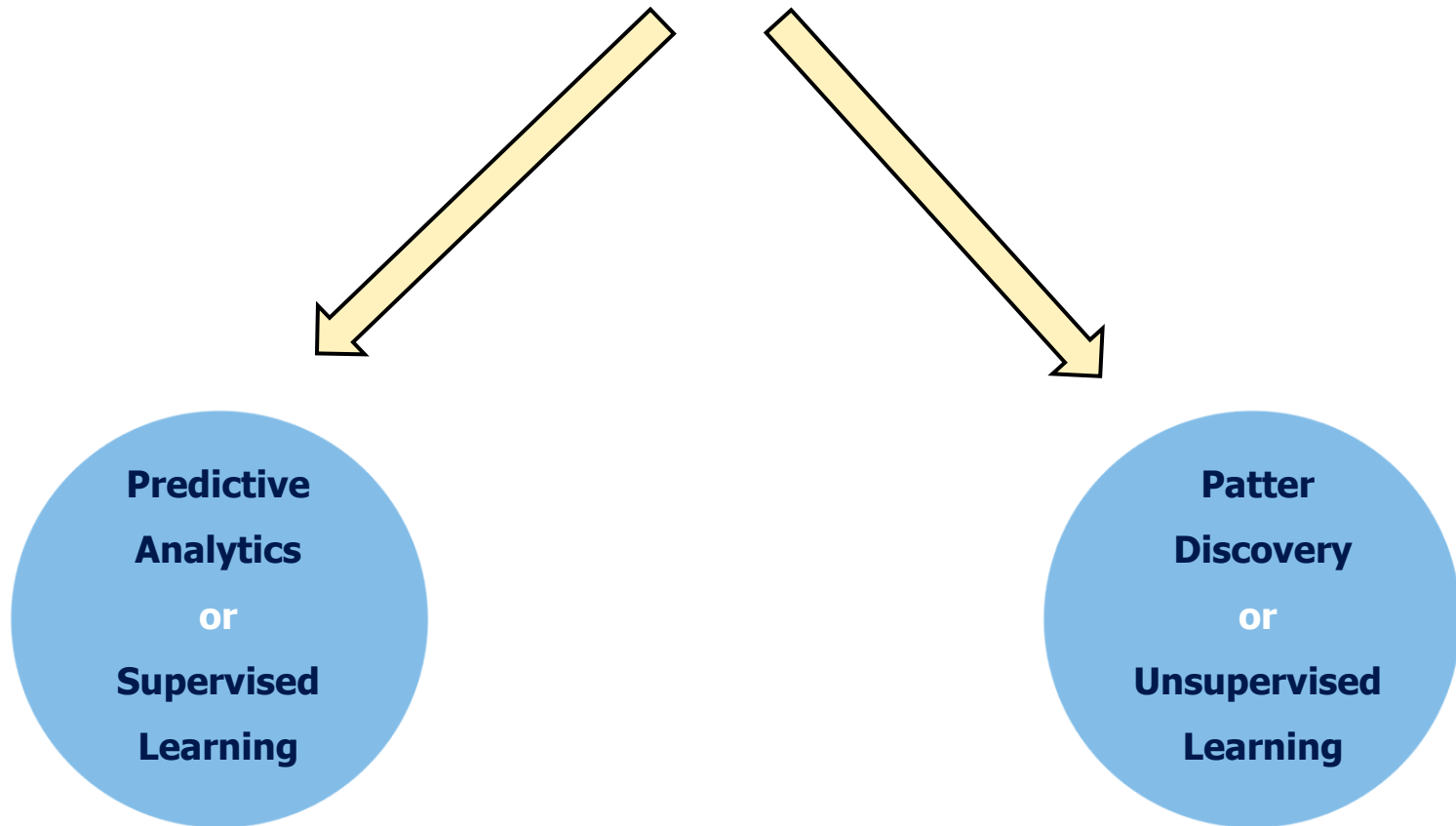
Applying machine learning to real-life problems can be difficult. It needs expertise, it is time consuming and costly. Research in AML helps automatically streamline the process using:

Automated Feature Engineering -- > Automated Model Selection -- > Automated Hyperparameter Tuning

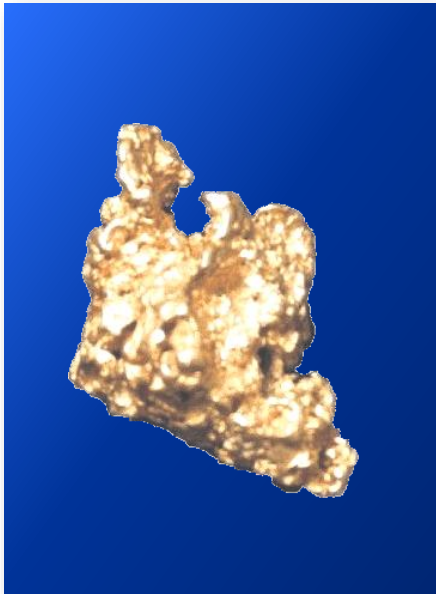
Self-taught Learning

The basic idea is to first learn a feature extractor from a large set of unsupervised learning data, extract features of a small set of labeled data, and then learn a classifier based on the features from the supervised learning data (Raina et al. from Stanford, 2007)

Data Mining/ Machine Learning Techniques



About Pattern Discovery/Unsupervised Learning



“...the discovery of interesting, unexpected, or valuable structures in large data sets.”

– David Hand

Pattern Discovery/ Unsupervised Learning Applications



Data reduction



Profiling

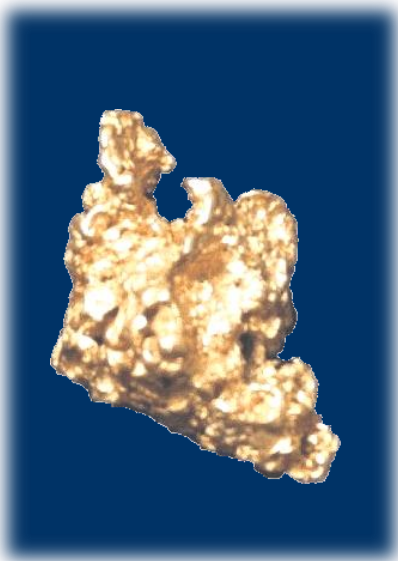


Market basket analysis



Sequence analysis

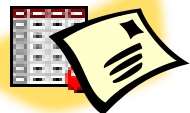
About Predictive Modeling/ Supervised Learning



“Most of the big payoff [in data mining] has been in predictive modeling.”

– Herb Edelstein

Predictive Modeling/ Supervised Learning Applications



Database marketing



Financial risk management



Fraud detection



Process monitoring

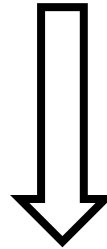


Pattern detection

Forecasting Vs Predictive Analytics/ Supervised Learning

Organization	Forecasting	Predictive Analytics
e - Commerce	Forecast the demand for each stock keeping unit (SKU) for the next six months.	Predict which customers are likely to respond to a promotion.
Telco	Forecast how many new customers will sign contract the next quarter.	Predict which customers are likely to churn (change provider).
Bank	Forecast the amount of deposits in the next three months.	Predict which customers that ask for a loan will repay their debt.
Insurance	Forecast the amount of money the customers will claim the next twelve months.	Predict which customer's claims are more likely to be fraudulent.

SAS Visual Data Mining & Machine Learning



Decision Support Tool for DM/ ML Projects

Predictive Analytics

Decision Trees

Logistic Regression

Neural Networks

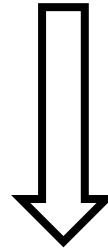
Pattern Discovery

Cluster Analysis

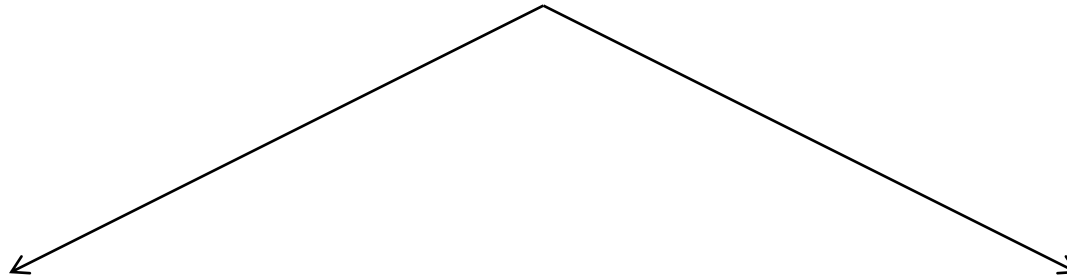
Association Analysis

Sequence Analysis

SAS Forecast Server/ SAS Visual Forecasting



Decision Support Tool for Forecasting Projects



Econometric Methods

Regression & Dynamic Regression

ARIMAX

Structural Models

Time Series Methods

Exponential Smoothing

ARIMA

Intermittent Demand

The Analytical Workflow

