



Boosting Your Data Analysis Power with Python

Part 2

Ray Chen

raychen_email@yahoo.com

Agenda

Predictive Analytics

Machine Learning Algorithms

Demo



Predictive Analytics

Predictive analytics is the process of extracting information from historical data to make predictions about future outcomes or trends.

- Defect Prediction: Identify potential defects or quality issues in products or processes.
- Maintenance Optimization: Optimize maintenance activities and minimize equipment downtime.

Predictive Analytics

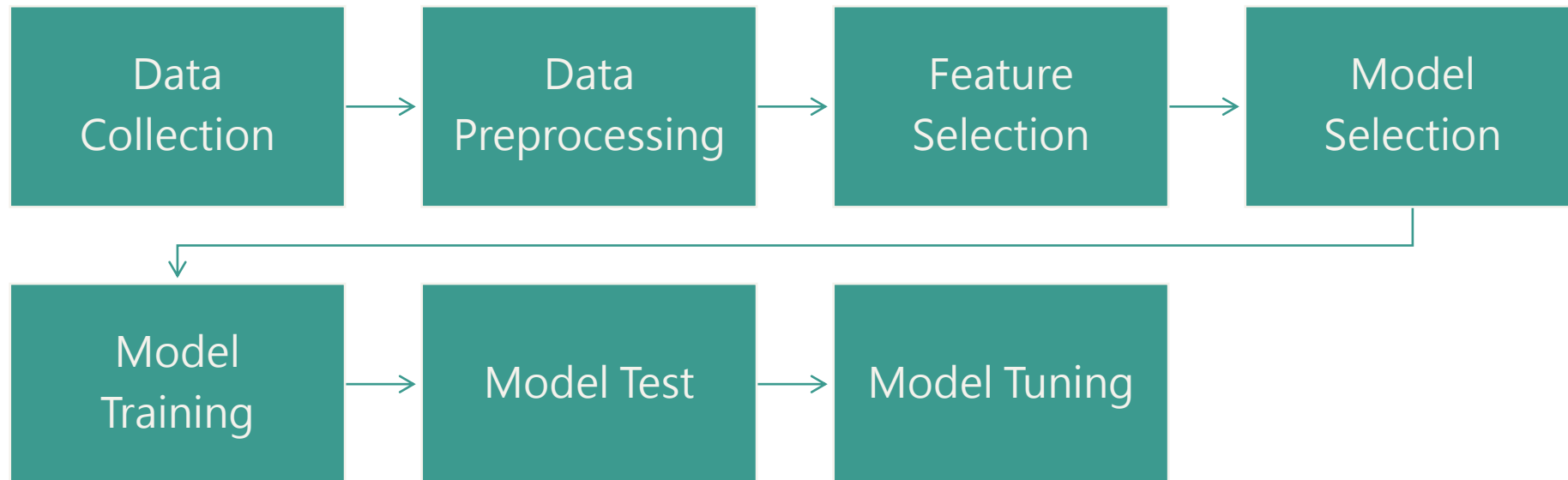
- **Supplier Quality Management:** Predicting the performance and reliability of suppliers, enabling proactive supplier management strategies.
- **Customer Satisfaction Prediction:** Predict customer satisfaction levels based on various factors such as customer feedback, purchase data, and demographic information, allowing organizations to take proactive measures to improve customer experiences and loyalty.

Machine Learning Algorithm

A machine learning algorithm is a computational procedure that utilizes statistical techniques to automatically learn from historical data and then make predictions on unseen data.

- Regression (Supervised Learning)
- Classification (Supervised Learning)
- Clustering (Unsupervised Learning)
- Pattern (Both)

Machine Learning Procedure



Models

You will see these in the demo

- CART (Classification and Regression Tree)
- Random Forest
- Gradient Boosting

CART

CART is a decision tree algorithm that recursively partitions the values of input features (split data) to create a tree-like model.

It can be used for both classification and regression tasks.

Random Forest

Random Forest is an ensemble learning method that builds an ensemble of decision trees by training each tree on a random subset of the training data (bootstrap samples) and considering only a random subset of the input features at each split.

- Reduce overfitting
- Robust to noisy or irrelevant features

Gradient Boosting

Gradient boosting is a machine learning technique that combines the predictions of multiple weak models (decision trees) to create a strong predictive model. It is an ensemble learning method where each subsequent model is built to correct the mistakes made by the previous models. Gradient boosting is based on the principle of gradient descent optimization.

Model Performance

- Model selection
- Cross validation
- Model tuning

Example

Plastic strength is influenced by many process inputs such as injection molding temperature, injection molding pressure, plastic formula, and type of injection molding machine. A predictive analytics model would identify solutions for maximizing the strength

- 1 response - strength
- 21 predictors – include machine settings, plastic materials, and injection molding machines



Thank You