Data Analytics with MATLAB

MathWorks India
What is Data Analytics?

“Data analytics solutions allow firms to discover, optimize, and deploy predictive models by analyzing data sources to improve business outcomes.”

Forrester Research

Challenges

• Overloaded by data
• Competition for better decision-making
• Big Data buzz, but missing solution to make sense of it (analytics) and use it (deployment)

Big Data Borat
@BigDataBorat

In Data Science, 80% of time spent prepare data, 20% of time spent complain about need for prepare data.

9:47 PM - 26 Feb 2013

291 RETWEETS 77 FAVORITES
Case Study: RITA Data Analytics

Goal:
- Create a data analytics tool, to get business intelligence from RITA flight dataset. Data available from 1987 till date

Challenges:
- Huge volume of Data (~ 30 GB)
- Selective data extraction and data mining
- Predictive analytics based on data
- End users need to be isolated form the technicalities

Solution:
- Create an interactive application, which
  - Showcases critical statistics from data
  - Performs data mining and machine learning
  - Allows user to perform predictive analytics
  - Creates detailed explanatory reports, giving relevant intelligence
Key take-aways

• Unified end to end platform for Data Analytics
  • Parallel processing to handle large data and intensive calculations
  • In-built machine learning algorithms for predictive analytics
  • Support for building, deploying and integrating analytics
Data Analytics Workflow
Tesco (Required Solution)
Tesco – Data Analytics Requirement

Data
- Temperature History
- Weather Forecast
- Sales History

Analytics Models
- Price Elasticity
- Sales uplift

Enterprise Systems
- Promotions management system
Tesco – Data Analytics Solution and Results

Data
- Temperature History
- Weather Forecast
- Sales History

Analytics Models
- Price Elasticity

Business Results
- £30m Less waste
- 30% Fewer shortages
- £50m Less Inventory

Automate
Tesco (Results Achieved)
## Integrated Platform for Data Analytics

<table>
<thead>
<tr>
<th>Challenges</th>
<th>MATLAB Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handling Large Data</td>
<td>Distributed arrays and Parallel Computing</td>
</tr>
<tr>
<td>Extract value from data</td>
<td>Built-in Statistics and Machine learning algorithms</td>
</tr>
<tr>
<td>Time to deploy &amp; integrate</td>
<td>Ease of deployment and leveraging enterprise</td>
</tr>
<tr>
<td>Technology risk</td>
<td>High-quality libraries and support</td>
</tr>
</tbody>
</table>
A scalable solution for big data computation!
Parallel and Distributed Computing

**Multicore Desktop**

12 Local Workers

Simulink, Blocksets, and Other Toolboxes

**Computer Cluster**

MATLAB Distributed Computing Server

Workers

Scheduler
Distributed Arrays: A PCT & MDCS Paradigm

Distributed Arrays

Multiple Flat Files

Combine results
Predictive Analytics - Overview

**Type of Learning**

- **Unsupervised Learning**
  - Group and interpret data based only on input data

- **Supervised Learning**
  - Develop predictive model based on both input and output data

**Categories of Algorithms**

- Clustering
- Classification
- Regression

Machine Learning
Predictive Analytics - Unsupervised Learning
Clustering with MATLAB

Group and interpret data based only on input data

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering
    - K-means, Fuzzy K-means
    - Hierarchical
    - Neural Network
    - Gaussian Mixture
Supervised Learning

Classification with MATLAB

Unsupervised Learning

Machine Learning

Supervised Learning

Classification

Develop predictive model based on both input and output data

- Decision Tree
- Ensemble Method
- Neural Network
- Support Vector Machine
Supervised Learning
Regression with MATLAB

Supervised Learning

Unsupervised Learning

Machine Learning

Develop predictive model based on both input and output data

Regression

Linear

Non-linear

Non-parametric
Key take-aways

- Unified end to end platform for Data Analytics
  - Parallel processing to handle large data and intensive calculations
  - In-built machine learning algorithms for predictive analytics
  - Support for building, deploying and integrating analytics