



Enterprise Computing Solutions - Education Services

## TRAINING OFFERING

---

**You can reach us at:**

Arrow ECS B.V., Kromme Schaft 5, 3991 AR Houten, The Netherlands

Email: [education.ecs.nl@arrow.com](mailto:education.ecs.nl@arrow.com)  
Phone: +31 20 582 6109

<b>CODE:</b>	<b>LENGTH:</b>	<b>PRICE:</b>
SPL_SFAADS	4.48 Hours (0.56 days)	€1,500.00

## Description

This 13.5 hour course is for users who want to attain operational intelligence level 4, (business insights) and covers implementing analytics and data science projects using Splunk's statistics, machine learning, built-in and custom visualization capabilities.

## Objectives

- Analytics Framework
- Exploratory Data Analysis
- Regression for Prediction
- Cleaning and Preprocessing and Feature Extraction
- Algorithms, Preprocessing and Feature Extraction
- Clustering Data
- Detecting Anomalies
- Forecasting
- Classification

## Prerequisites

- Splunk Fundamentals 1
- Splunk Fundamentals 2
- Splunk Fundamentals 3
- *or equivalent Splunk experience*

## Programme

### Module 1 – Analytics Workflow

- Define terms related to analytics and data science
- Define the analytics workflow
- Describe common usage scenarios
- Navigate Splunk Machine Learning Toolkit

### Module 2 – Exploratory Data Analysis

- Describe the purpose of data exploration
- Identify SPL commands for data exploration
- Split data for testing and training using the sample command

### Module 3 – Predict Numeric Fields with Regression

- Differentiate predictions from estimates
- Identify prediction algorithms and assumptions
- Describe the fit and apply commands
- Model numeric predictions in the MLTK and Splunk Enterprise
- Use the score command to evaluate models

### Module 4 – Clean and Preprocess the Data

- Define preprocessing and describe its purpose

- Describe algorithms that preprocess data for use in models
  - Use FieldSelector to choose relevant fields
  - Use PCA and ICA to reduce dimensionality
  - Normalize data with StandardScaler and RobustScaler
  - Preprocess text using Imputer, and NLP, TF-IDF, HashingVectorizer and the cluster command

#### Module 5 – Cluster Data

- Define Clustering
- Identify clustering methods, algorithms, and use cases
- Use Smart Clustering Assistant to cluster data
- Evaluate clusters using silhouette score
- Validate cluster coherence
- Describe clustering best practices

#### Module 6 – Anomaly Detection

- Define anomaly detection and outliers
- Identify anomaly detection use cases
- Use Splunk Machine Learning Toolkit Smart Outlier Assistant
- Detect anomalies using the Density Function algorithm
- Optimize anomaly detection with the Local Outlier Factor
- View results with the Distribution Plot visualization

#### Module 7 – Estimation and Prediction

- Differentiate predictions from forecasts
- Use the Smart Forecasting Assistant
- Use the StateSpaceForecast algorithm
- Forecast multivariate data
- Account for periodicity in each time series

#### Module 8 – Classification

- Define key classification terms
- Use classification algorithms
  - AutoPrediction
  - LogisticRegression
  - SVM (Support Vector Machines)
  - RandomForestClassifier
- Evaluate classifier tradeoffs
- Evaluate results of multiple algorithms

### **Session Dates**

On request. Please [contact us](#)

### **Additional Information**

This training is also available as onsite training. Please contact us to find out more.