

# Syllabus

## 강사 안건이

Contents	세부항목	소요시간	Key Word
Introduction to Data Analytics	Data Mining 이란 무엇인가? 데이터 분석의 목적 Python 환경 설정	4	가상 환경 관리
Data Loading from AWS(S3)	Data Loading from AWS Multi-Processing using Multi-core	3	AWS(S3) Multi-processing
Data Preprocessing with Python	10 Min Pandas Data Loading Data Cleaning Feature Extraction 데이터 실습	2	Pandas Domain
Basic of Data Analytics	Bias vs Variance Overfitting vs Underfitting Loss Function (Gradient Descent) K-fold Cross Validation	3	Trade-off Model Validation
Regression Problem Regularized Linear Models	Regression 이란 무엇인가? Regression 의 Loss Function Regression 지표 해석 데이터 실습	4	Linear Regression Ridge LASSO ElasticNet
Classification Problem Ensemble Learning Gradient Boosting Machine Family	Classification 이란 무엇인가? Classification 의 Loss Function Tree 계열 모델 설명(Ensemble) Tree 계열 지표 해석 데이터 실습	8	Decision Tree Ensemble Random Forest GBM, XGboost LightGBM
eXplainable Method For High Complexity Model	Global vs Local LIME / SHAP 결과 해석 기법 및 Plotting 데이터 실습	5	LIME SHAP Summary Plot Dependence Plot
Clustering & Dimensionality Reduction	Clustering 이란 무엇인가? Distance 계산 및 종류 Dimensionality Reduction 데이터 실습	6	Spectral HDBSCAN PCA T-SNE