|  |  |  |
| --- | --- | --- |
| **내용** | **관련 자료** | **비고** |
| **파이썬 기초 문법** | - |  |
| 기초 문법 | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| 파일 열기 | 03\_Python\_기초정리  02\_File Open Practice |  |
| 기본 분석환경 세팅 | 03\_Python\_기초정리 |  |
| Numpy | 03\_Python\_기초정리 |  |
| **EDA** | - |  |
| EDA | 04\_EDA\_시각화 | 기본적인 템플릿 정도 |
| **전처리** | - |  |
| 자료형 변환 | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| Aggregation, groupby, apply | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| Stack, unstack, melt, pivot | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| Concat, join, merge | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| 시간날짜 데이터 다루기 | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| 결측값 처리 | 03\_Python\_기초정리 |  |
| 중복값 처리 | 03\_Python\_기초정리 |  |
| 이상치 처리 | 03\_Python\_기초정리 |  |
| Dictionary, Array, List, Set 관련 | 03\_Python\_기초정리 |  |
| MultiIndex | 01\_Pandas Cookbook  03\_Python\_기초정리 |  |
| Dummy화 | 03\_Python\_기초정리 |  |
| Rolling, Shift | 03\_Python\_기초정리 |  |
| 오버샘플링, 언더샘플링 | 07\_기계학습\_1  패캠 자료 봐야함 ㅗㅗㅗㅗㅗㅗㅗㅗㅗㅗ |  |
| **시각화** | 04\_EDA+시각화 | Seaborn부터 시도 |
|  | 01\_Pandas Cookbook |  |
| **통계분석** | - |  |
| 빈도분석/분포분석 (기초통계량 Set) | 06\_통계분석\_1 |  |
| t-test | 05\_CDA | 평균값에 대한 검정 |
| ANOVA | 05\_CDA |  |
| 두 변수간 Correlation coefficient | 05\_CDA |  |
| 직접확률계산법 | X | 비율에 대한 검정 |
| F분포법 | X | 비율에 대한 검정 |
| 교차분석(Chi-square) | 05\_CDA | 분할표의 검정 |
| Fisher의 직접확률 검정 | 06\_통계분석\_1 |
| 맥네마의 검정 | 06\_통계분석\_1 |
| 잔차 분석 | X |
| 상관 분석 | 06\_통계분석\_1 | 변수들간 상관관계 강도 |
| 단순회귀분석 X1, Y1 | 06\_통계분석\_1 | 변수간 선형/비선형 인과관계의 형태와 강도 추출 |
| 다중회귀분석 Xk, Y1 | 06\_통계분석\_1 |
| 다항회귀 X1^n, Y1 | 06\_통계분석\_1 |
| *규제 적용한 회귀* | - |
| 정규화선형회귀-Ridge 회귀 | 06\_통계분석\_1 |
| 정규화선형회귀-Lasso 회귀 | 06\_통계분석\_1 |
| 정규화선형회귀-Elastic Net 회귀 | 06\_통계분석\_1 |
| 일반화선형회귀-로지스틱 Xk, Y(0,1) | 06\_통계분석\_1 |
| 판별분석 |  |
| 다차원척도법MDS-계량적(구간/비율) |  | 대상 분류 후 다차원 공간에 배치 |
| 다차원척도법-비계량적(순서척도) |  |
| 대응분석 |  | 패턴 분류(설문답변) |
| 시계열분석 |  |  |
| 연관성분석 |  |  |
| 장바구니분석 |  |  |
| **기계학습** |  |  |
| train\_test\_split | 07\_기계학습\_1 |  |
| Scaling(정규화, 표준화) | 07\_기계학습\_1 |  |
| *회귀* | - |  |
| 선형회귀 | 06\_통계분석\_1 |  |
| 의사결정나무(DT) | 06\_통계분석\_1 |  |
| Random Forest Reg | 06\_통계분석\_1 |  |
| Gradient Boosting Reg | 06\_통계분석\_1 |  |
| XGB Reg | 06\_통계분석\_1 |  |
| LGB Reg | 06\_통계분석\_1 |  |
| 혼합모델(결과 더하는 거) | 06\_통계분석\_1 |  |
| 스태킹 앙상블 | 필요하면 파머 완벽 가이드 | p.369 |
|  | ㅗㅗㅗㅗㅗ 패캠 자료보기 |  |
| *분류* | - |  |
| 로지스틱 휘귀 | 06\_통계분석\_1 |  |
| 나이브 베이즈분류 | ㅗㅗㅗㅗㅗ 패캠 자료 |  |
| K-Nearest Neighbor Classifier |  |  |
| SVM |  |  |
| 의사결정나무 | 07\_기계학습\_1 |  |
| 앙상블 – Voting | 07\_기계학습\_1 |  |
| 앙상블 – Bagging - RF Classifier | 07\_기계학습\_1 |  |
| 앙상블 – Boosting – GB, XGB, LGB | 07\_기계학습\_1 |  |
| 스태킹 앙상블 | 필요하면 파머 완벽 가이드 | p.275 |
| *군집화* | - |  |
| K-Means Clustering | 07\_기계학습\_2 |  |
| 평균 이동 | 07\_기계학습\_2 |  |
| GMM | 07\_기계학습\_2 |  |
| DBSCAN | 07\_기계학습\_2 |  |
| 계층 군집 분석(Hierarchical Cluster) |  |  |
|  | ㅗㅗㅗㅗㅗㅗ 패캠 자료 |  |
| *차원축소* | - |  |
| 요인분석 | ㅗㅗㅗㅗㅗㅗㅗㅗ | 핵심 요인 선별 |
| PCA 주성분분석 | 07\_기계학습\_1 | 핵심 요인 선별 |
| LDA 선형 판별분석 | 07\_기계학습\_1 |  |
| SVD | 07\_기계학습\_1 |  |
| NMF | 07\_기계학습\_1 |  |
|  |  |  |
| 연관규칙분석 | ㅗㅗㅗㅗㅗㅗ |  |
| 사회연결망 분석(SNA) | ㅗㅗㅗㅗㅗㅗ |  |
|  |  |  |
| *최적화기법(단계적 변수선택)*  *Feature selection(sklearn에도 있던데)* | 6-1 회귀 뒤로 붙이자 |  |
| 전진선택법(Forward Selection) |  |  |
| 후진제거법(Backward elimination) |  |  |
| 단계선택법(Stepwise method) |  |  |
| **모델 성능평가** | - |  |
| 회귀(R2, RMSE, MSE, MAE, MAPE, MSPE, MSAE 공식, 함수) | 07\_기계학습\_1  ㅗㅗㅗㅗㅗㅗㅗㅗ |  |
| 분류(정확도, 특이도, 민감도, 정밀도, 재현율, F1, ROC, AUC) | 07\_기계학습\_1  ㅗㅗㅗㅗㅗㅗㅗㅗㅗㅗ |  |
| 군집화(실루엣지수, SSE, Dunn Index) | 07\_기계학습\_2 |  |
| **Neural Network** | - |  |
|  | ㅗㅗㅗㅗ패캠 보기 |  |
| ANN | 7-2 |  |
| CNN |  |  |
| DNN |  |  |
| RNN |  |  |
| **텍스트 마이닝** | - |  |
|  | 7-2 |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

