

ADP 시험분석

데싸라면

ADP 시험 분석

Advanced Data Analytics Professional

◎ 데이터분석 전문가의 직무

직무	세부내용
데이터 기획	비즈니스 목표 달성을 위해 내부 업무 프로세스를 기반으로 다양한 분석기회를 발굴하여 분석의 목표를 정의하고, 분석대상 도출 및 분석 결과 활용 시나리오를 정의하여 분석과제를 체계화 및 구체화하는 빅데이터분석과제 정의, 분석로드맵 수립, 성과 관리 등을 수행한다.
데이터분석	분석에 대한 요건을 구체적으로 도출하고, 분석과정을 설계하고, 요건을 실무담당자와 합의 하는 요건정의, 모델링, 검증 및 테스트, 적용 등을 수행한다.
데이터 시각화	다양한 데이터들을 대상으로 어떤 요소를 시각화 해야 하는지 정보 구조를 분석하며 어떤 형태의 시각화 모델이 적합한지 시각화에 대한 요건을 정의하고 시나리오를 개발하는 시각화 기획, 모델링, 디자인, 구축, 배포 및 유지보수 등을 수행한다.

21회 기출 문제 분석

1. 머신 러닝 (50점)

학생 성적에 관한 데이터셋 제공 약 350행짜리 소규모 데이터.

1-1. 시각화 포함 탐색적 자료분석(EDA)

1-2. 결측치 식별하고 결측치를 예측하는 두 가지 방법 정도를 쓰고, 선택한 이유를 설명.

1-3. 범주형 변수 인코딩이 필요한 경우를 식별하고, 변환을 적용하시오. 선택한 이유를 설명.

1.4. 데이터 분할 방법을 2가지 쓰고 적절한 데이터 분할을 적용. 선택한 이유 설명.

1.5. svm, xgboost, randomforest 3개의 알고리즘 공통점을 쓰고 이 예측 분석에 적합한 알고리즘인지 설명.

1.6. 세 가지 모델 모두 모델링 해보고 가장 적합한 알고리즘 선택하고 이유 설명. 한계점 설명하고 보완 가능한 부분 설명.
현업에서 사용시 주의할 점 등에 대해 기술.

21회 기출 문제 분석

2. 통계분석 (50점)

1. 연속형 독립변수 여러개의 소규모 데이터. (총 29점)

변수명은 순서대로 $x_1 \sim x_{10}$, 의미는 없는 데이터

1-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 rmse 구하시오

1-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오.

alpha 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 rmse를 계산

1-3. 라쏘 회귀로 2-2과 동일한 문제

2. 독립변수 하나 종속변수 하나 소규모 데이터 다항회귀(12점)

다항 회귀를 3차까지 적용하고 각 차수별 데이터포인트 스캐터 플롯과 계수와 기울기 선을 그리세요.

3. ANOVA분석 (9점)

변수 3개(하나는 범주형 변수/ 나머지 두 개는 수치형 연속변수)

이원분산분석을 수행하고 통계표를 작성하시오.

22회 기출 문제 분석

1. 기계학습(50점)

1.데이터 탐색(당뇨 데이터 세트, 데이터에 헤더가 없음, 대신 시험지에 변수명 제공)

1.1.1 탐색적 데이터 분석 수행하시오(시각화 포함)

1.1.2 이상치 처리하시오

1.1.3 앞선 두 단계에서 얻은 향후 분석시 고려사항 작성

2. 클래스 불균형을 처리하시오

2.1 업 샘플링 과정 설명하고 결과 작성

2.2 언더 샘플링 과정 설명하고 결과 작성

2.2 둘 중 선택하고 이유 설명

3. 모델링 하시오

3.1 최소 3개 이상 알고리즘 제시하고 정확도 측면의 모델 1개와 속도 측면의 모델 1개를 꼭 구현(총 2개 이상)

3.2 모델 비교하고 결과 설명

3.3 속도 개선을 위한 차원 축소 설명하고 수행, 예측 성능과 속도 비교하고 결과 작성

22회 기출 문제 분석

2. 통계분석 (50점)

1. 금속 성분 함유량 변수 1개. (1열 데이터) 제품에 금속 재질 함유량의 분산이 1.3을 넘으면 불량이라고 보고 제조사별로 차이가 난다고 제보를 받는다. **분산에 대해 검정을 수행하시오.**

1.1 연구가설과 귀무가설 작성 / 1.2 양측 검정 / 1.3 검정통계량, 가설 채택

2.Lot별 불량 제품 수량 데이터. lot 번호와 불량제품수 두 개의 열. 각 lot별 200개에 대한 불량제품 수.

2.1 p관리도에 따라 관리중심선(center line), 관리 상한선, 하한선 구하시오

2.2 관리도 시각화 하시오

3.표에 제품 1,2를 만드는데 사용되는 재료 a b c 컬럼 있고 재료에 따라 최종 만들어지는 제품 두 개에 대한 수량 있다.

제품 수량을 최대로 뽑으면서 수익이 최적이 되도록 작성하시오.(10점)

4. 상품 a와 b가 있을 때 다음과 같은 구매 패턴이 있다고 함. aa bb bbbb aa aaa bb bbb aa bb a b

4.1구매하는 패턴으로 봐서 두 상품이 연관이 있는지 가설 세우고 검정하시오

4.2 연구가설 귀무가설 세우시오

4.3 가설 채택하시오

ADP 시험 유형1.

[통계분석 + 시각화]

[15회, 2019-12-14]

데이터 처리 및 통계 분석

- timestamp 처리 / date 기준 데이터 병합
- hh:mm, A/B/C/D/E, 전력 사용량 데이터
- yyyyymmdd, 평균 기온

1. 아래 형태의 데이터 생성 : 3개 데이터를 date기준으로 병합 필요
→ yyyy-mm / A / B / C / D / E / 사용량
2. 요일 변수 생성 및 A/B/C/D/E별 평균 사용량 출력, 그래프 출력
3. 요일간 사용량 분석을 수행하고 가장 차이가 있는 요일 도출

2019

[17회, 2020-06-21]

통계 분석 (설문데이터 분석)

→ (사전에 역문항들에 대한 처리 필요)

1. 그룹별 평균, 표준편차, 왜도, 첨도 산출

시각화 및 시계열 분석 (코로나 데이터) (20점)

1. 전체 인구대비 누적 사망률이 가장 높은 5개 국가 추출 후, 국가별 일일확진자, 누적확진자, 일일사망자, 누적사망자 시계열 그래프 출력
2. 위험지수 생성 및 해석
3. 시계열 분석 및 예측 모델 생성

2020

[21회, 2021-06-09]

2. 연속형 독립변수 여러개의 소규모 데이터.

변수명은 순서대로 x1~x10 이라 의미 없음

2-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 rmse 구하시오

2-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오. alpha 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 rmse를 계산

2-3. 라쏘 회귀로 2-2과 동일한 문제

3. 독립변수 하나 종속변수 하나 소규모 데이터.

다항 회귀를 3차까지 적용하고 각 차수별 데이터포인트 스캐터 플롯과 계수와 기울기 선을 그리라고 함(12점) .

4. ANOVA 분석(9점)

변수 3개

하나는 abcde 각각을 값으로 갖는 범주형 변수

나머지 두 개는 수치형 연속변수

이원분산분석을 수행하고 통계표를 작성하시오.

2021

ADP 시험 유형 2.

[시계열분석 + 시각화]

[18회, 2020-09-19]

시계열 분석

- 데이터 구성 : Year / month / amount

정상성 확인 : 평균과 분산 일정 + 근거 & 해석 (10점)

ARIMA 모델 3가지 제시 (10점)

최적 모델 선택 및 근거 서술 (15점)

최종 예측 후 실제 결과와 비교 평가, 평가 방법 사용 이유 제시 (15점)

[19회, 2020-12-13]

시계열 분석

- 날짜 및 주가 수익률 데이터 (Time-Series 변환 전 데이터 제공)

데이터 로드, 정상성/이분산성 검증

정상성 파악 근거에 따른 고정시계열 여부 파악

SARIMA 분석 및 최적 모형 파라미터 선택

잔차 그래프 출력

2020

[20회, 2021-03-27]

- 날짜 및 온도 데이터

1. 시계열분석(DATA : 온도예측) (총 50점)

1-1 : 데이터 전처리 (10점)

1-2 : 위 데이터를 RandomForest로 검증하고 분석하기(15점)

1-3 : 위 데이터를 SVM(서포트 벡터 머신)으로 검증하고 분석하기(15점)

1-4 최적의 모델 선택하기

2021

ADP 시험 유형 3.

[데이터마이닝 / 머신러닝 + 시각화]

[15회, 2019-12-14]

제조 생산 데이터 분석

데이터 탐색 : EDA

데이터 전처리 : 변수 선택(VIF), 파생변수 생성,

데이터 분할(train/validation/test(20%))

로지스틱 분석 : 분류1 을 판단 하는 모델 생성

(종속변수는 총 7개 분류, 분류1 외의 값은 0으로

치환), confusion matrix 해석

로지스틱 분석 외 3개 이상 분류 모델 생성 및 결

과 해석 : SVM 필수 포함, Precision/Sensitivity

결과 출력

위 모델 중 최고 모델을 선택하여 최적의 군집 개

수를 선택하고 클러스터링 수행 : F-1 Score 출

력

[17회, 2020-06-21]

집값 예측 문제

EDA

모델 생성

데이터 분할

교호작용을 고려한 다중 선형 회귀 수행

3가지 분류 모델 생성 및 비교, 좋은 모델 선택

[18회, 2020-09-19]

고객 등급 예측 모형

EDA & 결측값 채우기

파생변수 3개 생성 & 이유 작성

Train-Test 분할(7:3) / SOM 군집분석 / 정오분류표

분류분석 4가지

[19회, 2020-12-13]

고객 이탈 예측

20개의 x와 1개의 y를 가지는 데이터로 y는 0과 1을 가짐

EDA 및 전처리

분류모델 3개 적용 및 Confusion Matrix 출력

3개 모델 앙상블 모형 생성 후 예측값 저장

[21회, 2021-06-05]

데이터 마이닝을 사용하여 중등 학교 학생 성과 예측날짜 및 주가 수
익률 데이터 (Time-Series 변환 전 데이터 제공)

1-1. 시각화 포함 탐색적 자료분석

1-2. 결측치 식별하고 결측치를 예측하는 두 가지 방법 정도를 쓰고,

선택한 이유를 설명1-3. 범주형 변수 인코딩이 필요한 경우를 식별하

고, 변환을 적용하시오. 선택한 이유를 설명1.4. 데이터 분할 방법을 2

가지 쓰고 적절한 데이터 분할을 적용. 선택한 이유 설명.1.5. svm, xg

boost, randomforest 3개의 알고리즘 공통점을 쓰고 이 예측 분석에

적합한 알고리즘인지 설명.1.6. 3모델 모두 모델링 해보고 가장 적합

한 알고리즘 선택하고 이유 설명. 한계점 설명하고 보완 가능한 부분

설명. 현업에서 사용시 주의할 점 등에 대해 기술.

2019

2020

2021

ADP 시험 유형 4. [텍스트 마이닝]

[12회, 2019-03-30]

뉴스기사 로딩 및 제공된 긍정/부정 어휘를 통한 감성분석
위에서 구한 긍정부정 score를 통해 N개의 그룹으로 클러스터링

[18회, 2020-09-19]

텍스트 마이닝 (문제 상세 파악 안됨)
명사 추출 & 불용어 처리
빈도 막대 그래프

ADP 실기시험 13회부터 17회 사이에는
텍스트 마이닝 과목이 출제되지 않았습니다.

2019

2020

ADP 준비 방법

ADP 실기 코드북 ☆ 🔒 📄

파일 수정 보기 삽입 서식 도구 부가기능 도움말 22분 전에 마지막으로 수정했...

100% 일반 텍스트 Nanum Got... 11 B I U A

2 1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

🔥 DataFrame 컬럼명 변경

```
df.rename( { 'old1':'new1', 'old2':'new2' }, axis='columns' )
df.columns = [ 'new1', 'new2' ]
df.columns = df.columns.str.replace("기존문자", "대체할 문자") # 컬럼명 특정문자 대체
```

🔥 DataFrame 컬럼 타입 확인

```
df.dtypes
```

🔥 DataFrame 컬럼 타입 변경

```
df['컬럼명'] = df['컬럼명'].astype('타입')

# 모든 컬럼에 대해 가능한 가장 적절한 dtypes로 변경
df = df.convert_dtypes()
```

2. 행/열 선택/추가/삭제

🔥 행/열 선택

행의 index 또는 컬럼명을 사용하는 방법

```
df['컬럼명'] # 하나의 컬럼을 Series 형식으로 출력
df[['컬럼명1', '컬럼명2']] # 여러 개의 컬럼은 Data Frame 형식으로 출력
df.loc[[ 행 index, 컬럼명 list]] # 행과 열을 지정하여 출력
# 모든 행 또는 모든 열을 출력할 때는 : 을 대신 사용
```

