


K-Digital Training

최종 프로젝트

프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅
팀 명	2조
문서 제목	가명처리관련정보문서

Version	1.0
Date	2021-06-10

팀원	강 길웅 (팀장)
	곽 지선
	김 진현
	박 주희
	배 효린
	임 종배
	하 주연

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10

CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 SK 인포섹 K-Digital Training 교육 과정 중 최종 프로젝트 2조의 팀원들의 자산입니다. 교육 관련 관계자 및 K-Digital Training 교육과정 2조의 팀원들의 서면 허락없이 사용되거나, 재 가공 될 수 없습니다.

문서 정보 / 수정 내역


Filename	k-digital-training_최종프로젝트_기획서.docx
원안작성자	강길웅
수정작업자	곽지선, 김진현, 박주희, 배효린, 임종배, 하주연

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2021-06-10	강길웅	1.0	최초 작성	

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10

목 차

1	가명처리	4
1.1	가명처리 수준 정의표.....	4
1.2	가명처리 검토 결과.....	6
2	가명처리 로직	7
2.1	가명처리 과정 설명.....	7
2.2	의사코드.....	7
3	데이터 적정성 평가 결과	11

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10

1 가명처리


1.1 가명처리 수준 정의표

< 가명처리 수준 정의표 >

순번	테이블명	컬럼명	식별	처리수준
1	회원정보	회원번호	가명식별자	회원번호 + SALT 를 SHA256 암호화 후 기존 회원번호 삭제
2	회원정보	이름	식별정보	삭제
3	회원정보	ID	식별정보	삭제
4	회원정보	Password	식별정보	삭제
5	회원정보	생년월일	식별정보	10세 단위로 범주화 (20세 미만 하단코딩, 80세이상 상단코딩)
6	회원정보	성별	식별정보	- (다른 속성값을 가명처리하여 위험을 낮추도록 조치)
7	회원정보	주소 (우편번호)	식별가능정보	앞에서부터 3자리를 남긴 후 일반화
8	회원정보	핸드폰번호	식별정보	삭제
9	회원정보	이메일	식별정보	삭제
10	회원정보	신장	식별가능정보	상단 코딩 처리 (남자 : 200cm 이상의 경우 → 200cm 여자 : 190cm 이상의 경우 → 190cm)
11	회원정보	체중	식별가능정보	상단 코딩 처리

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10


				(200kg 이상의 경우 → 200kg)
12	회원정보	보험여부	식별가능정보 (민감정보)	해부화 (Y = [1, 12, 3, 14, 5, 16, 7, 18, 9], N = [11, 2, 13, 4, 15, 6, 17, 8, 19])
13	건강정보	건강정보번호	-	삭제
14	건강정보	혈압	식별가능정보	-
15	건강정보	걸음수	식별가능정보	-
16	건강정보	심박수	식별가능정보	-
17	건강정보	혈당	식별가능정보	-
18	건강정보	체온	식별가능정보	-
19	건강정보	측정 날짜	식별가능정보	-
20	결제정보	결제정보번호	-	삭제
21	결제정보	카드사	식별가능정보	삭제
22	결제정보	카드번호	식별정보	삭제
23	결제정보	카드CVC번호	식별정보	삭제
24	결제정보	은행명	식별가능정보	삭제
25	결제정보	계좌번호	식별정보	삭제
26	결제정보	결제금액	식별가능정보	앞에서부터 2자리를 남긴 후 자릿수 만큼 0을 입력하여 랜덤라운딩 (156400 → 150000)
27	결제정보	결제날짜	식별가능정보	일반화 (YYYYMM)

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10

1.2 가명처리 검토 결과

< 가명처리 검토 결과 보고서 >

가명정보 활용목적	헬스케어 데이터 예측 모델 구축을 위한 과학적 연구 수행		
가명처리 대상항목 (원본 데이터)	회원번호, 이름, ID, Password, 생년월일, 성별, 주소(우편번호), 핸드폰번호, 이메일, 신장, 체중, 보험여부, 건강정보번호, 혈압, 걸음 수, 심박 수, 혈당, 체온, 측정 날짜, 결제정보번호, 카드사, 카드번호, 카드CVC번호, 은행명, 계좌번호, 결제금액, 결제날짜		
처리(제공) 환경 검토	처리 환경	클라우드 상에서 원본DB와 가명DB로 별도로 분리하여 가명처리한 데이터를 저장하며 가명정보를 제공	
	제공 받는 자의 환경	내부	
	제공 받는 자의 보호수준	AWS Parameter Store 사용하여 salt값을 암호화하여 저장, IAM 계정과 security group으로 권한을 분리하여 접근 제한 등	
항목별 위험도 분석	<ul style="list-style-type: none">- 회원번호는 식별자이므로 회원번호와 salt값을 결합하여 SHA256으로 암호화 후 기존 회원번호를 삭제 처리(가명식별자로 대체)- 이름, ID, Password, 핸드폰번호, 이메일, 카드번호, 카드CVC번호, 계좌번호는 식별정보이므로 삭제 처리- 주소(우편번호)는 식별가능정보이지만 지역을 기반으로 한 분석이 필요하여 앞에서부터 3자리를 남긴 후 일반화 처리- 체중은 식별 가능 정보이므로 특이치 기준을 200으로 잡고 상단코딩 처리 (ex: 200kg 이상인 경우 200kg)- 생년월일은 식별정보이므로 10세단위로 범주화하고 20세 미만은 하단코딩, 80세 이상은 상단코딩 처리- 신장은 식별가능정보이므로 남/여 특이치 기준을 정하여 상단코딩 처리 (ex: 여: 190cm이상 → 190cm, 남: 200cm이상 → 200cm)- 성별은 식별정보이지만 분석의 목적에 필요하므로 다른 속성값을 가명처리하여 위험을 낮추도록 조치- 보험여부는 식별가능정보(민감정보)이므로 Y=[1, 12, 3, 14, 5, 16, 7, 18, 9] 와 N=[11, 2, 13, 4, 15, 6, 17, 8, 19] 로 해부화 처리- 카드사, 은행명은 식별가능정보이며 분석의 목적에 반드시 필요하지 않으므로 삭제 처리- 결제금액은 식별가능정보이므로 앞에서부터 2자리를 남긴 후 자릿수만큼 0을 입력하여 랜덤라운딩 처리 (ex: 156400 → 150000)- 결제날짜는 식별가능정보이므로 일반화(YYYYMM) 처리		

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10

2 가명처리 로직

2.1 가명처리 과정 설명

1. 원본 데이터베이스 RDS 에서 테이블 내용 읽어옴
2. Select 쿼리문을 통해 회원 ID 추출
3. 파라미터 스토어에서 Salt 값 추출
4. [회원 ID + Salt] 를 SHA 256 암호화
5. 컬럼 각각 비식별화 처리
6. 가명처리 데이터베이스 테이블에 데이터 삽입
 - a. if [회원 ID + Salt] 를 SHA 256 암호화한 값이 가명처리 디비의 테이블에 있으면
-> 가명처리 디비에 있는 그 회원의 데이터를 업데이트 (삭제 후 신규 값 삽입)
 - b. 없으면
-> 신규 값 삽입

2.2 의사코드

```

import random
from Cryptodome.Hash import SHA256

def birth(b):
    If b < 20:
        B = 10
    else if b >= 80:
        B = 80
    else :
        B의 첫번째 자리의 수는 그대로 고정
        B의 나머지 뒷자리 수는 0으로 변경
    return b;

def topcoding(height, sex):
    If sex == 여자 :
```



```
if height >= 190 :  
    height = 190
```

else:

```
if height >= 200 :  
    height = 200
```

```
return height;
```

def post_code_generalization(x):

```
    x를 리스트화( list(x) )  
    앞에서부터 3자리는 그대로 남긴 후, 뒤의 4번째, 5번째자리의 값을 제거  
    (list[ : 2 ] 슬라이싱 이용)  
    x를 문자열화  
    return x
```

def pay_date_generalization(x):

```
    x를 리스트화( list(x) )  
    앞에서부터 6자리는 그대로 남긴 후, 뒤의 7번째, 8번째자리의 값을 제거  
    (list[ : 5 ] 슬라이싱 이용)  
    x를 문자열화  
    return x
```

def price_rounding(x):

```
    x를 리스트화( list(x) )  
    x의 길이 계산  
    앞에서부터 2자리는 그대로 남긴 후, 뒤의 나머지 값을 자릿수만큼 0으로 변경  
    x를 문자열화  
    return x
```

원본 데이터 RDS에 있는 데이터 테이블 불러오기()

if 불러온 테이블 값이 NULL이 아니면 :

```
    회원 ID 컬럼 값 추출. (Select 쿼리문 이용)  
    Parameter Store를 이용해 SALT 추출.
```




암호화된 회원ID = (회원ID + SALT)를 SHA256으로 암호화 처리

암호화된 회원 ID = SHA256.new(SALT)

암호화된 회원 ID.update(회원ID.encode('utf-8'))

암호화된 회원 ID.hexdigest()

식별자 삭제 처리

회원 ID, 이름, ID, Password, 핸드폰번호, 이메일,

건강정보번호, 결제정보번호, 카드사, 카드번호, 카드CVC번호, 은행명, 계좌번호 컬럼
삭제

(ex. Drop(columns = ['이름'] 이용)

앞에서 부터 3자리를 남긴 후 일반화

우편번호 = post_code_generalization(우편번호)

상단 코딩 처리

topcoding(신장, 성별)

생년월일 범주화

birth(생년월일)

해부화 (보험 여부를 특정 구간 값으로 변환)

y_list = [1, 12, 3, 14, 5, 16, 7, 18, 9]

n_list = [11, 2, 13, 4, 15, 6, 17, 8, 19]

for i in 보험 여부 컬럼 데이터 길이 :

if 보험 여부(i) == 'Y' :

보험 여부(i) = random.choice(y_list)

else :

보험 여부(i) = random.choice(n_list)

앞에서 부터 2자리를 남긴 후 자릿수만큼 0을 입력

결제금액 = price_rounding(결제금액)

앞에서 부터 6자리를 남긴 후 일반화

결제날짜 = pay_date_generalization(결제날짜)

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10

가명처리한 데이터를 가명처리 RDS에 삽입 (삽입시 **회원 ID**는
암호화된 회원 ID로 삽입) (Insert 쿼리문 사용)

완료되었습니다 성공 메시지 출력

else :

테이블을 불러오지 못했습니다 오류 메시지 출력
RDS에 있는 테이블 다시 불러오기()

 K-Digital Training 최종 프로젝트	가명처리관련정보문서		
	프로젝트 명	클라우드 기반 헬스케어 시스템 아키텍처 설계 & 컨설팅	
	팀 명	2조	
	Confidential Restricted	Version 1.0	2021-06-10


3 데이터 적정성 평가 결과

가명처리 적정성검토 결과서

평가 번호	P-2021-001					
평가위원 정보	성명	김경하	소속	제이앤시큐리티	직위	대표
평가 일자	시작일자 일시 : 2021 년 06 월 05 일 14 시 00 분 종료일자 일시 : 2021 년 06 월 05 일 15 시 00 분					
평가 결과	<input checked="" type="checkbox"/> 적정 <input type="checkbox"/> 부적정					
평가 의견	<ul style="list-style-type: none"> - 가명처리 목적, 이용 환경, 데이터 특성 및 보호조치 수준을 고려해 볼 때, 현재의 가명처리 수준은 적절한 것으로 판단됨 - 데이터 구조(컬럼 등)와 가명처리 방법/수준이 동일한 경우, 주기적·반복적 가명처리 시 추가적인 적정성검토는 불필요할 것으로 사료됨(단, 데이터구조 또는 가명처리 방법/수준의 변경이 발생할 경우에는 가명처리 적정성 검토 수행 필요) - 또한, 데이터 누적 등에 따라 재식별 가능성이 증가할 수 있으므로, 정기적(연 1회 등)으로 가명처리 적정성 검토를 수행할 것을 권고함 					
최초평가 해당여부	<input checked="" type="checkbox"/> 주기적·반복적 가명처리의 최초평가에 해당 <input type="checkbox"/> 해당없음					

위와 같이 적정성 평가 결과를 통지합니다.

2021년 6월 5일

서 명 란	
-------------	--