개인정보 비식별화 개요 및 비식별 기술 개요

NIA 한국정보화진흥원

목 차

▣ 교육 개요

개인정보 비식별화 개요 및 비식별 기술 개요

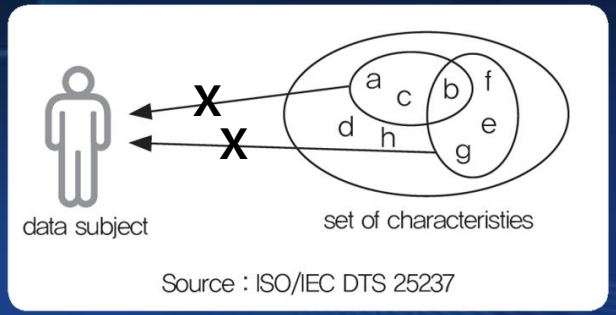
빅데이터 활용과 관련하여 여러 제약사항에 대하여 알아보고 이를 해결하기 위한 방안 고찰

▣ 목 차

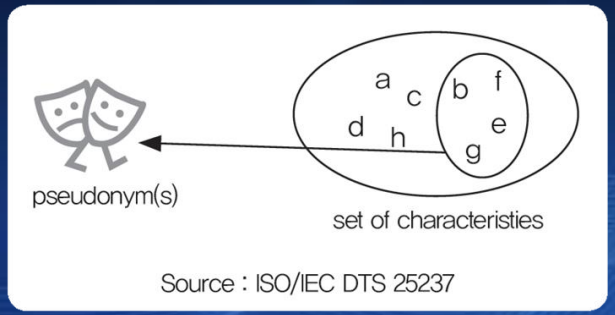
1. 비식별화 개요

▣ 비식별화 개념

– 데이터 내에 개인을 식별할 수 있는 정보가 있는 경우, 이의 일부 또는 전부를 삭제, 또는 일부를 속성 정보로 대체 처리함으로써 다른 정보와 결합하여도 특정 개인을 식별하기 어렵도록 하는 조치



Source : ISO/IEC DTS 25237



Source : ISO/IEC DTS 25237

Anonymization (익명화) -> Generalization (일반화) and Perturbation (섭동, 변경)

1. 비식별화 개요

▣ 비식별화 대상 및 기준

o 적용대상 : 그 자체로 개인을 식별할 수 있는 정보 및 해당 정보만으로 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보들을 대상으로 함

o 적용시기 : 빅데이터 수집·활용의 전(全) 단계에서 개인정보가 식별되는 경우 혹은 이후 정보의 추가 가공 등을 통하여 개인이 식별되는 경우 등

※ 예시) ① 개인 정보의 수집 및 저장 시, ② 개인 정보가 포함되어 있을 수 있는 데이터의 활용 시,

③ 다른 기관(정보)과의 정보 공유 시, ④ 기관내의 서로 다른 부서간의 정보 공유 시

1. 비식별화 개요

▣ 비식별화 대상 및 기준

그 자체로 개인 식별이 가능한 정보는 삭제

- 단, 수집 시에 개인정보에 대한 자체이용, 제3자 제공 등 활용에 대한 이용자 동의를 받았을 경우 비식별화 없이 활용 가능

다른 정보와 결합에 따른 재식별 위험 최소화

- 보유 개인정보의 분석을 위한 동의 등이 곤란한 경우 분석 목적을 달성할 수 있는 한도에서 비식별화 처리

정보가 식별 될 수 있는 리스크를 고려하여 사후관리 철저

- 주기적으로 재식별에 대한 리스크 검토 및 리스크 통제 가능한 매커니즘 확보
- 빅데이터 분석 등의 과정에서 불필요한 개인정보가 새로 생성되거나 비식별화 처리된 정보가 재식별화 된 경우에는 지체 없이(통상 5일이내) 삭제하거나 비식별화 처리 등

[붙임] 비식별화 적용 대상 예시

구분	주요 내용
① 그 자체로 개인을 식별할 수 있는 정보 (식별자)	• 쉽게 개인을 식별할 수 있는 정보 : 이름, 전화번호, 주소, 생년월일, 사진 등 • 고유식별정보 : 주민등록번호, 운전면허번호, 의료보험번호, 여권번호 등 • 생체정보 : 지문, 홍채, DNA 정보 등 • 기관, 단체 등의 이용자 계정 : 등록번호, 계좌번호, 이메일 주소 등 • 기타 유일 식별번호 : 군번, 사업자등록번호 특성(별명), 식별코드(아이디, 아이핀 값(cn, dn)) 등
② 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보 (준식별자)	• 개인특성 : 성별, 생년, 생일, 연령(나이), 국적, 고향, 거주지, 시군구명, 우편번호, 병역여부, 결혼 여부, 종교, 취미, 동호회·클럽, 흡연여부, 음주여부, 채식여부, 관심사항 등 • 신체 특성 : 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔, 신체검사 결과, 장애유형, 장애등급, 병명, 상병코드, 투약코드, 진료내역 등 • 신용 특성 : 세금 납부액, 신용등급, 기부금, 건강보험료 납부액, 소득분위, 의료급여자 등 • 경력 특징 : 학교명, 학과명, 학년, 성적, 학력, 직업, 직종, (전·현)직장명, 부서명, 직급, 자격증명, 경력 등 • 전자적 특성 : PC사양, 비밀번호, 비밀번호 질문/답변, 쿠키정보, 접속일시, 방문일시, 서비스 이용 기록, 위치정보, 접속로그, IP주소, MAC주소, HDD Serial 번호, CPU ID, 원격접속 여부, Proxy 설정여부, VPN 설정 여부, USB Serial 번호, Mainboard serial 번호, UUID, OS 버전, 기기 제조사, 모델명, 단말기 ID, 네트워크 국가 코드, SIM Card 정보 등 • 가족 특성 : 배우자, 자녀, 부모, 형제 여부, 가족정보, 법정대리인 정보 등 • 위치 특성 : GPS 데이터, RFID 리더 접속 기록, 특정 시점 센싱기록, 인터넷 접속, 핸드폰 사용기록 사진 등

※ **민감정보 :** 개인의 사생활을 드러낼수 있는 속성
(비식별화 기법들에서 값을 보존하는 경우에 해당하는 데이터 분석시 주로 측정되는 대상 속성)

1. 비식별화 개요

▣ 비식별화 용어 이해

o 공개된 개인정보 : 이용자(정보주체) 및 정당한 권한이 있는 자에 의해 일반 공중에게 공개된 부호·문자·음성·음향·영상 등의 정보로서 생존하는 개인을 식별할 수 있거나 다른 정보와 쉽게 결합하여 개인을 식별 할 수 있는 정보

o 이용내역 정보 : 정보통신서비스와 관련하여 이용자가 해당 서비스를 이용하는 과정에서 자동으로 발생되는 인터넷 접속정보파일, 거래 기록 등의 정보로서 생존하는 개인을 식별할 수 있거나 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보

o 민감 정보 : 특정한 개인의 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 정보

o 정보 조합 ·분석 시스템 : 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 조합, 분석, 처리하여 정보를 생성하는 시스템

o 생성된 개인정보 : 정보 조합·분석·처리시스템 운용을 통해 생성된 정보로 개인을 식별할 수 있는 정보 및 다른 정보와 결합하여 개인을 식별할 수 있는 정보

o 비식별화 : 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 결합하여도 특정 개인을 식별할 수 없도록 하는 조치

o 재식별화 : 비식별화된 정보를 조합, 분석 또는 처리하는 과정에서 개인정보가 재 생성되는 것

2. 비식별화 기술 개요

▣ 비식별화 기술 소개

식별자 처리(일반적 기법)를 통한 식별방지+ 프라이버시 모델 기반 추론방지

처리기법	내용
가명처리 (Pseudonymisation)	• 개인 식별이 가능한 데이터에 대하여 직접적으로 식별 할 수 없는 다른 값으로 대체
총계처리 (Aggregation)	• 개인정보에 대하여 통계값(전체 혹은 부분)을 적용하여 특정 개인을 판단할 수 없도록 함
데이터 값 삭제 (Data Reduction)	• 개인정보 식별이 가능한 특정 데이터 값 삭제
범주화 (Data Suppression)	• 단일 식별 정보를 해당 그룹의 대표값으로 변환(범주화)하거나 구간값으로 변환(범위화)하여 고유 정보 추적 및 식별 방지
데이터 마스킹 (Data Masking)	• 개인 식별 정보에 대하여 전체 또는 부분적으로 대체값(공백, '*', 노이즈 등)으로 변환

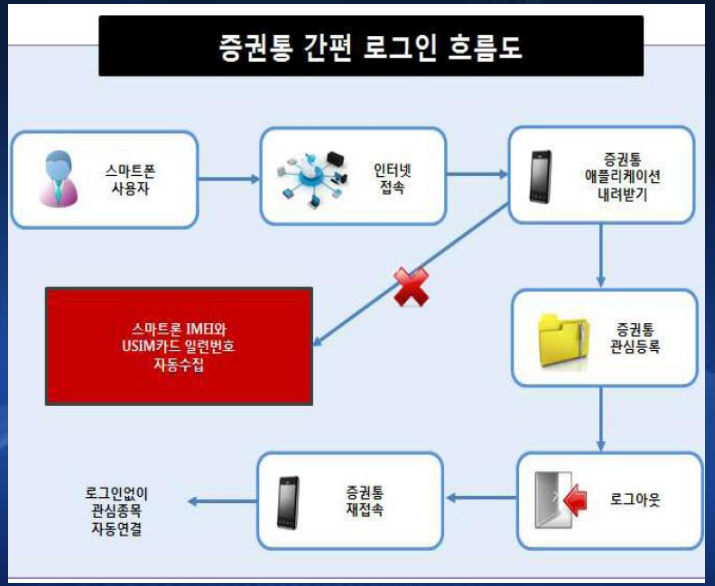
처리기법	내용
k-익명성 (k-anonymity)	• 특정인임을 추론할 수 있는지 여부를 검토, 일정 확률수준 이상 비식별 되도록 함
l-다양성 (l-diversity)	• 특정인 추론이 안된다고 해도 민감한 정보의 다양성을 높여 추론 가능성을 낮추는 기법
t-근접성 (t-closeness)	• l-다양성 뿐만 아니라, 민감한 정보의 분포를 낮추어 추론 가능성을 더욱 낮추는 기법

3. 비식별화 정책 필요성 – 1) 증권통 '간편 로그인'서비스

▣ 주요내용

o 주식시세정보를 제공하는 스마트폰 앱 "증권통"은 사용자들이 로그인 없이 사전에 등록한 관심종목을 볼 수 있도록 앱을 설계하기 위해 스마트폰 인증번호인 IMEI와 USIM 일련번호를 수집

※ 사용자들이 앱을 설치할 때 사용자의 스마트 폰으로부터 개인정보인 'IMEI(국제모바일단말기 인증번호)와 USIM 일련번호의 조합' 정보를 읽어 오거나, 'IMEI와 사용자 개인휴대전화번호의 조합' 정보를 읽어 와 서버에 저장한 다음, 사용자가 다시 접속하는 경우 서버에 저장된 사용자의 개인정보와 비교하여 사용자의 동일성을 식별하고 별도 로그인 없이 바로 사용자가 등록해 놓은 관심종목을 보여주도록 설계



3. 비식별화 정책 필요성 – 1) 증권통 '간편 로그인'서비스

▣ 추진 중단이유

o **IMEI나 USIM 일련번호를 개인정보로 보고 무단수집에 대해 법원 유죄판결**

o 결합의 가능성에 대하여 입수 가능성과는 상관없이 **결합만 가능하면 개인정보라고 판시**

- 증권통 관계자는 IMEI나 USIM 일련번호는 특정개인에게 부여된 부호가 아니라 특정기기, 특정카드 등에 부여된 번호이고, 다른 정보화 쉽게 결합하여 이용자가 알아보는 것이 불가능함으로 개인정보가 아니라고 주장

※ IMEI정보는 통신사가 가지고 있는 정보로 일반인은 접근이 불가

- 그러나 법원은, 개인정보보호법 상 개인정보의 정의에 결합의 용이성은 다른 정보의 '입수 가능성'을 전제하고 있지 않고, **결합의 용이성**을 말하고 있기 때문에 IMEI나 USIM 일련번호도 개인정보라고 판시

※ 기계적인 정보라 하더라도 특정 개인에게 부여되었음이 객관적으로 명백하고, 이러한 정보를 통하여 **개인이 식별될 가능성이 있다면 이를 개인정보로 본다**고 판단

3. 비식별화 정책 필요성 – 2) 개인신용평가 전문기업의 신용평가 개선모델

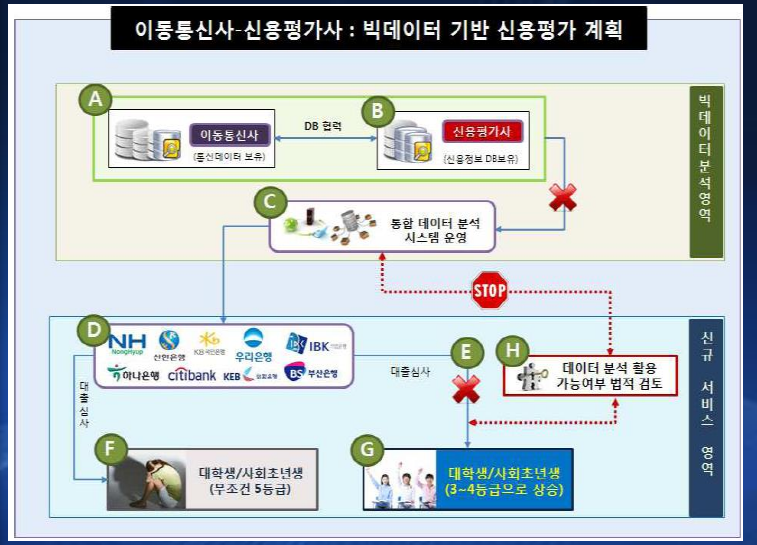
▣ 주요내용

o ○○개인신용평가 전문기업은 이동통신사의 결제패턴 등의 정보를 활용하여 신용등급이 낮거나 없는 대학생 및 사회 초년생에게 보다 좋은 대출 등급을 제공하기 위한 융합 신용평점 작업 추진

※ 자사의 개인정보데이터와 이동통신사의 통신 데이터의 매시업을 통해 통합 빅데이터 분석시스템을 운영하고, 이를 기반하여 특정 계층의 신용 평점 작업을 진행하여 고객사인 은행 등에 개인별 신용 평가 등급을 제공한다는 계획

※ 통신 사용량, 미수납 연체 등 정보를 활용하면 보다 높은 신용등급 부여가 가능하다는 아이디어

o 대출이 어려웠던 학생 및 일부 계층에서 사회 적응 가능성을 높이고 은행 및 대출 기관에서는 보다 신뢰성있는 데이터에 기반하여 대출 심사 가능



3. 비식별화 정책 필요성 – 2) 개인신용평가 전문기업의 신용평가 개선모델

▣ 추진 중단이유

o 신용평가회사 – 은행 – 통신사간 비식별화를 통한 정보 융복합이 개인정보보호법상 문제가 되는지 명확한 규정이 없어 적극적 사업 추진에 애로

※ 법무팀에서도 보수적인 의견을 제시

o 서로 다른 기관 간에 비식별 데이터 유통을 통해서 혁신적인 서비스를 만들어 낼 수 있으나, 비식별화에 대한 기준 미비 및 법적인 근거 미비로 사업 추진 보류

3. 비식별화 정책 필요성 – 3)사례 카드회사 빅데이터기반 타겟 마케팅

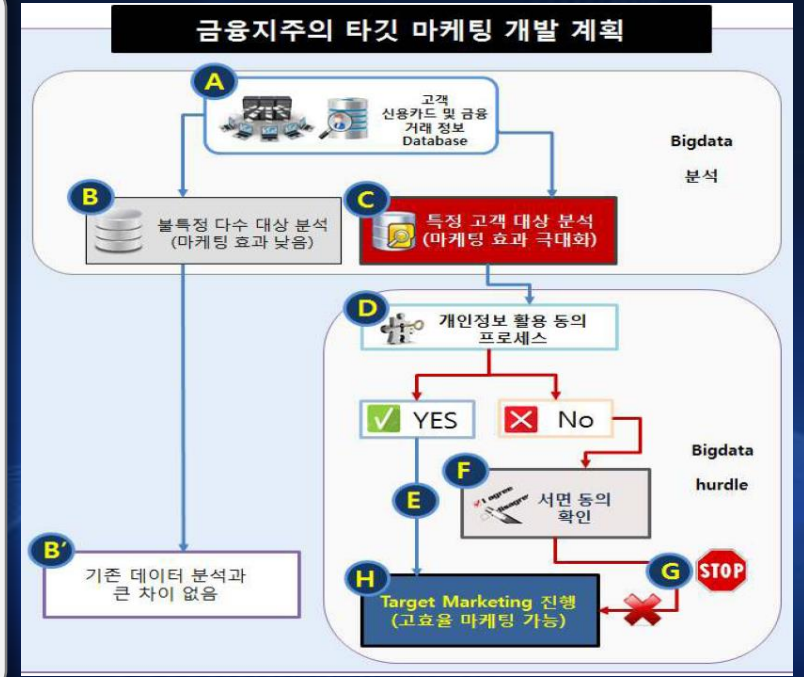
▣ 주요내용

o ○○카드사는 자사 카드의 이용고객 및 금융 거래 정보를 빅데이터 기술로 분석하여 특정 개인에 대한 타켓 마케팅을 전개하려고 계획

※ 불특정 다수를 대상으로 하는 현 마케팅 효과가 미미하여, 개인 식별이 가능한 데이터 분석을 통한 타겟 마케팅을 기획

o 이를 위해 모든 금융정보 및 거래 데이터를 데이터베이스화하고 특정 고객의 거래 패턴 및 지출 예측이 가능한 분석 시스템을 마련

※ 고객의 신용카드 사용 및 거래 정보, 은행 거래 데이터의 분석작업을 통해 마케팅 효과를 극대화할 수 있는 알고리즘 개발을 진행



3. 비식별화 정책 필요성 – 3)사례 카드회사 빅데이터기반 타겟 마케팅

▣ 추진 중단이유

o 금융감독원 감독규정에는 이들 타겟 마케팅을 위해서는 **고객 개별의 서면동의가 필요하다는 지침**을 주고 있어 장애물로 작용

o 새로운 마케팅을 전개하기 위해서는 **매번 서면 동의를 받기 위한 절차가 필요**하여 매우 번거롭고, 비용과 시간이 과다하게 소요된다는 점이 투자를 망설이게 하는 원인

※ 또한 통합된 고객정보가 없기 때문에 중복 마케팅 등 고객 불만 유발 가능성이 있다고 판단

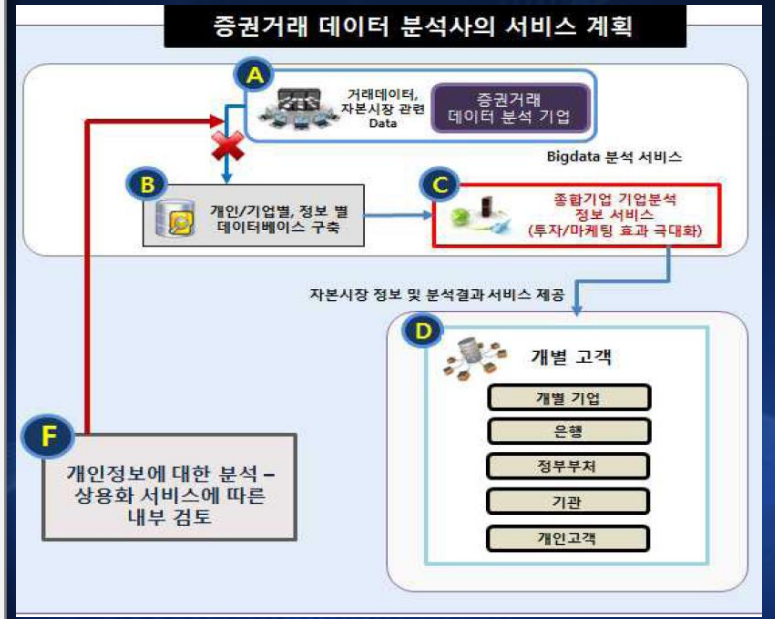
3. 비식별화 정책 필요성 – 4) 증권거래 데이터 분석

▣ 주요내용

o ○○자본시장 IT솔루션기업은 사내에서 보유하고 있는 증권, 주식 거래 등의 데이터를 매시업하여 보다 진보된 분석 정보를 투자 수요에 공급하여 새로운 수익을 창출하려는 시스템 구축 계획 마련

– 이를 위해 개인 거래, 기업 거래 및 각종 정형, 비정형을 모두 처리할 수 있는 데이터베이스를 구축하여 은행, 정부부처, 기관투자자, 개인 고객 등 에게 제공하여 새로운 수익을 창출하고자 함

※ 회사 내 각 부서에 요청하여 관련된 데이터를 제공 받아 투자성향, 추천 항목 등을 도출하는 빅데이터 분석을 시도



3. 비식별화 정책 필요성 – 4) 증권거래 데이터 분석

▣ 추진 중단이유

o **실데이터를 보유**하고 있는 회사 내 부서에서 개인 고객의 거래 정보를 분석할 경우 **고객 정보에 대해 유출 가능성을 우려**하여 사업 중단

※ 카드사 사태와 잇따른 개인정보 유출사고로 인해 **개인정보보호법**을 비롯해 **정보통신망법**, **신용정보보호법** 등에 개인정보 유출시 처벌조항과 과징금 등 관련 규제를 한층 강화

o 관련 부서의 책임자들에 대한 설득이 진행 중이나 법제화를 통한 명확한 책임 소재가 분명해질 때까지 프로젝트는 중단

3. 비식별화 정책 필요성 – 5) 신용평가시스템 개선

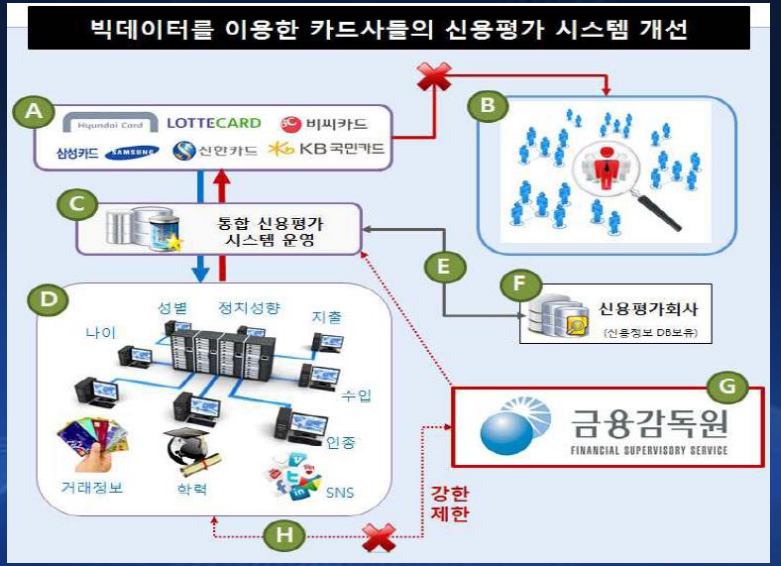
▣ 주요내용

o ○○카드사는 소셜데이터를 비롯하여 거래정보, 학력에 따른 거래정보, 정치성향 별 거래 정보 등 다양한 데이터 매시업을 통해 신용등급을 책정하는 알고리즘 개발 추진

※ 카드사 내부 거래 데이터는 물론 PG(Payment Gateway)사와 외부 신용평가 회사 데이터와의 매시업을 통해 보다 높은 예측력을 가지는 마케팅 기법을 도입하고자 함

o 이를 위해 카드사들은 신용평가회사의 데이터를 통합시키고 개인별 성향 데이터(나이, 성별, 거래정보, 학력, SNS활동, 인종, 수입, 지출 등)를 모두 통합하여 빅데이터 기반 기술을 통해 통합 신용평가 시스템 구축 기획

※ SNS 빅데이터를 이용한 개인신용평가 시스템은 수많은 SNS 사용기록을 분석 후 이를 지표화하는 것으로 제도권 금융기록을 기반으로 하는 개인신용평가시스템(Credit Scoring System)과 더불어 대출신청자들의 소셜 신용도까지 접목한 새로운 신용평가 시스템으로 활용 가능



3. 비식별화 정책 필요성 – 5) 신용평가시스템 개선

▣ 추진 중단이유

○ **'학력변수'의 이용에 대한 금융감독원의 강한 제한**이 중단의 가장 큰 이유

※ 학력변수 활용에 대해 금융감독원의 지적이 있어 SNS 및 기타 데이터 매시업 자체에 부담을 느끼고 사업을 중단

- 이로 인해 **내부 데이터 분석만 진행**하고 외부 데이터 및 감사에서 지적받을 만한 데이터는 모두 분석하지 않는 쪽으로 결정

3. 비식별화 정책 필요성 - 시사점

- **우리나라도 개인정보 주체의 권리는 강화하되 개인정보 처리자가 빅데이터와 같은 신산업 분야에서 개인정보를 투명하고 안전하게 활용할 수 있도록 지원 방안 필요**

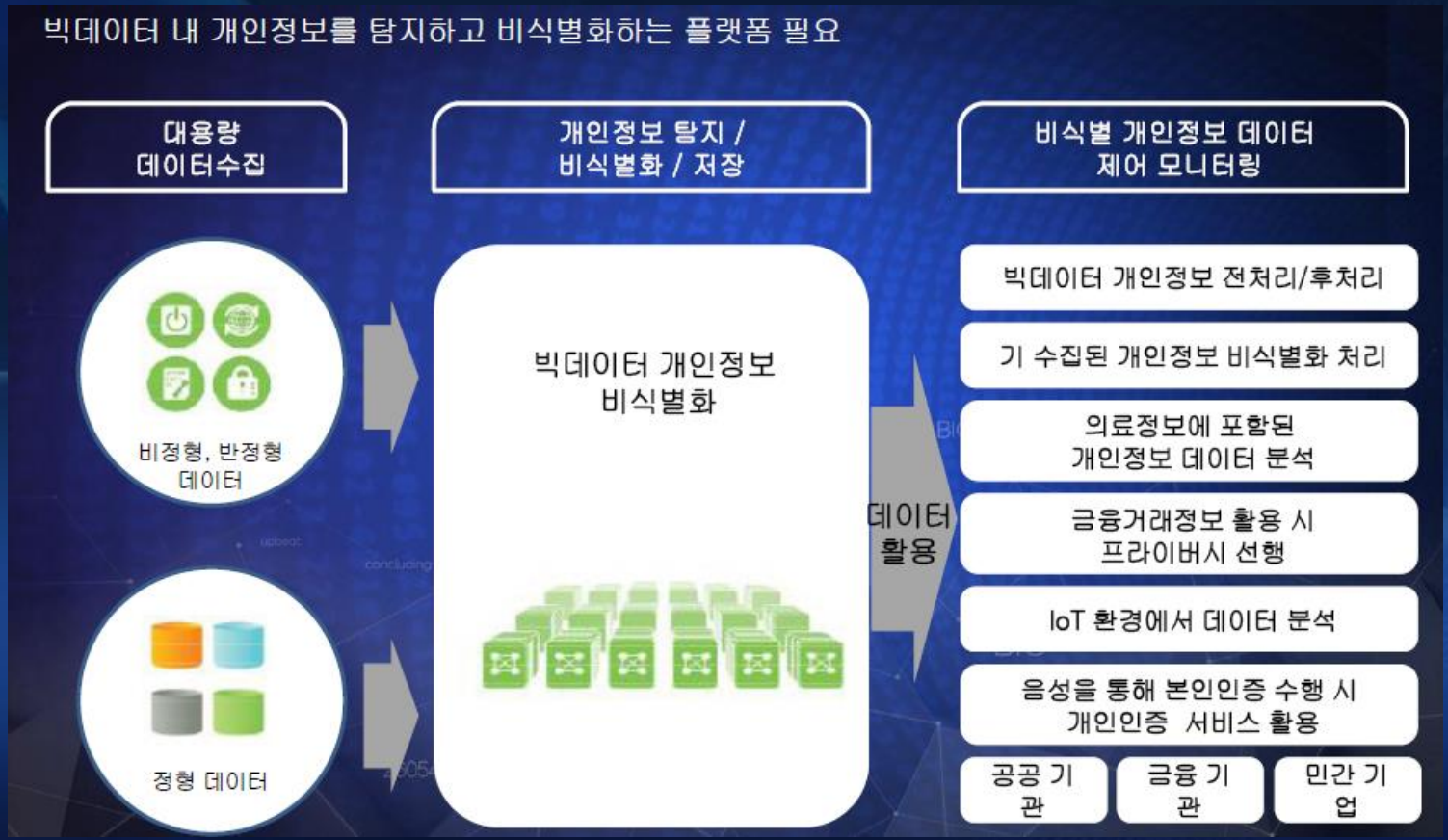
✓ **(제도개선)** 미래부·행자부(개인정보보호법)·방통위(정보통신망법)·금융위(신용정보법)·복지부 등 주요 관계부처 간 협업을 통해 균형있는 개인정보보호 관련 제도 개선 방안 마련 필요

✓ **(기술조치)** 법제도 개선과 함께 개인정보 비식별화 조치에 필요한 기술 개발 및 활용 안내서 마련, 관련 교육실시 등 병행 필요

✓ **(활용지원)** 빅데이터 유통 활용 관련 저해요소에 대한 상담 및 법률자문, 비식별화 등 기술적 컨설팅을 제공하는 클리어링 기능 강화

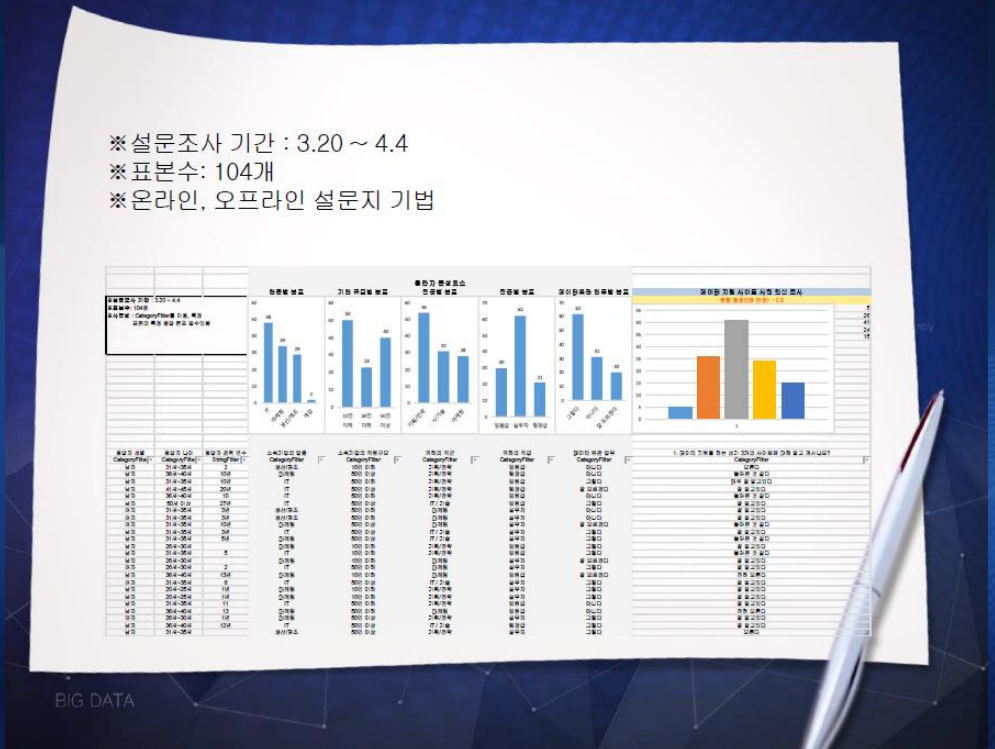
✓ **(공감대 형성)** 컨퍼런스·토론회 등을 통해 해외규제 비교분석, 현행 규제로 인한 불합리한 사례 공유, 홍보 등을 통해 규제개선 필요성에 관한 공감대 형성 중요

3. 비식별화 정책 필요성

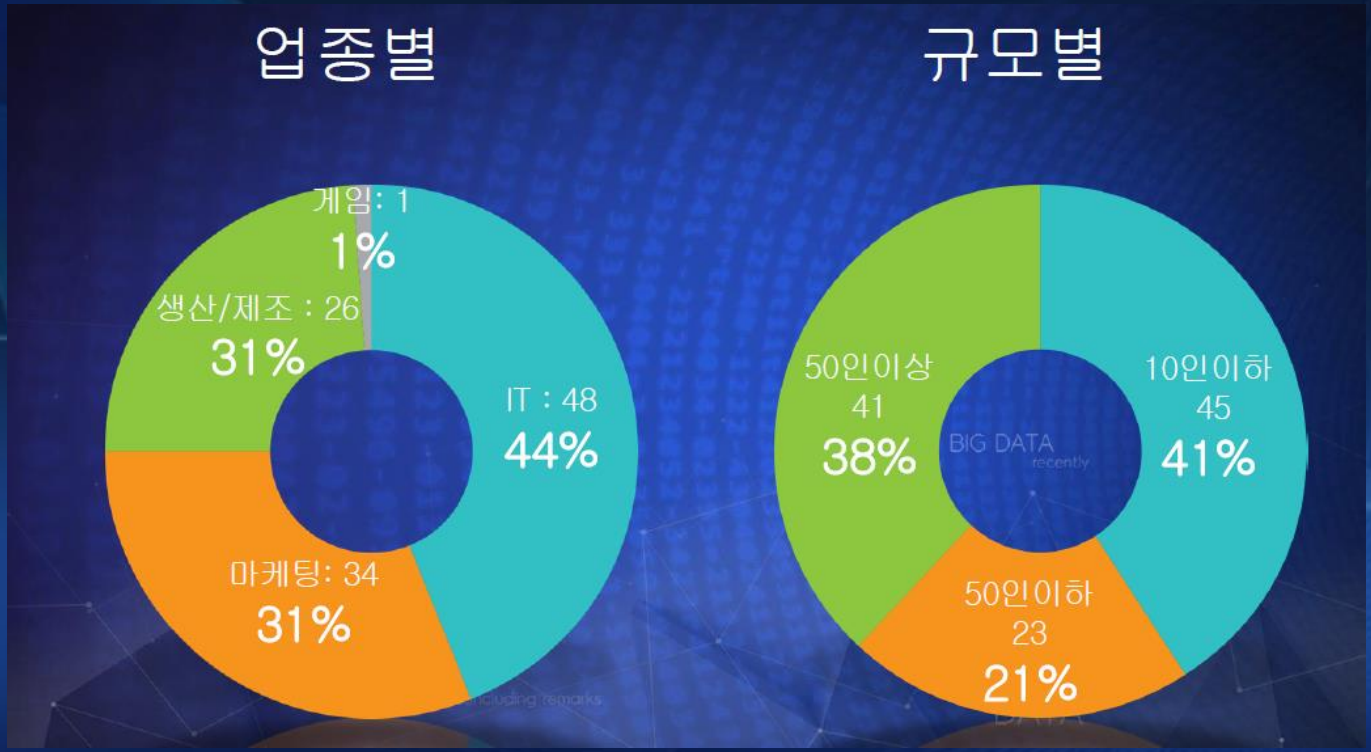

빅데이터 내 개인정보를 탐지하고 비식별화하는 플랫폼 필요
대용량 데이터수집
개인정보 탐지 / 비식별화 / 저장
비식별 개인정보 데이터 제어 모니터링
비정형, 반정형 데이터
정형 데이터
빅데이터 개인정보 비식별화
데이터 활용
빅데이터 개인정보 전처리/후처리
기 수집된 개인정보 비식별화 처리
의료정보에 포함된 개인정보 데이터 분석
금융거래정보 활용 시 프라이버시 선행
IoT 환경에서 데이터 분석
음성을 통해 본인인증 수행 시 개인인증 서비스 활용
공공 기관
금융 기관
민간 기업

4. SK 텔레콤 설문조사 (1)

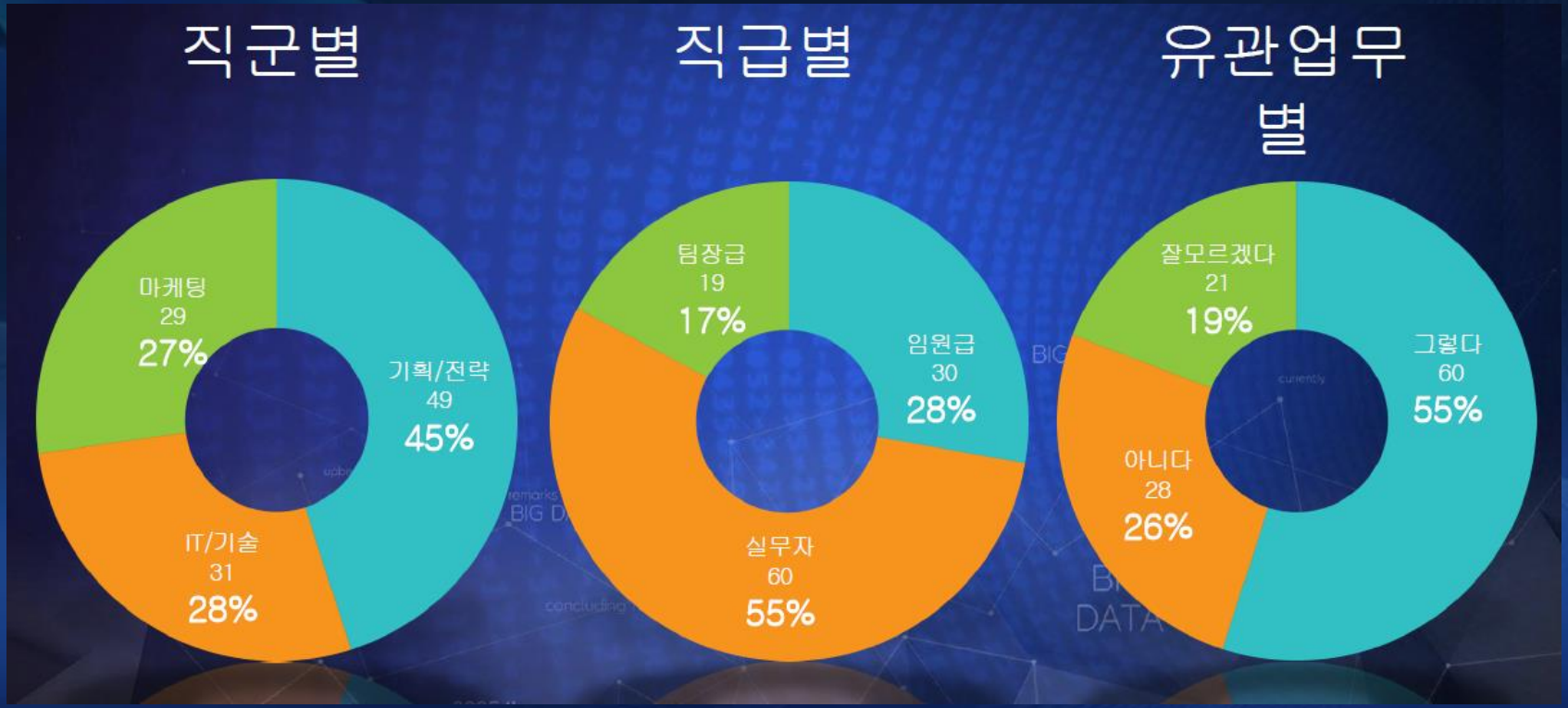
- 비식별화 조치 가이드라인 인지도 및 활용 기대수준 조사



4. SK 텔레콤 설문조사 (2)

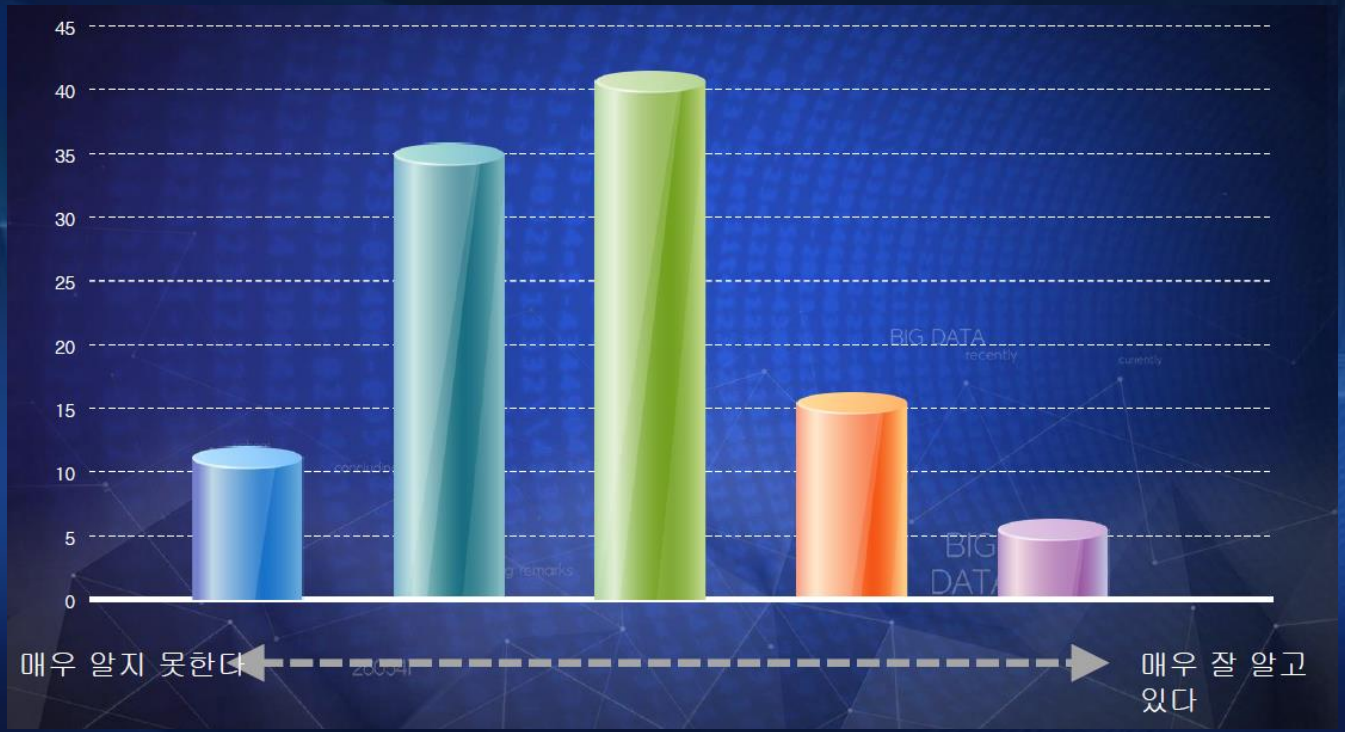


4. SK 텔레콤 설문조사 (3)



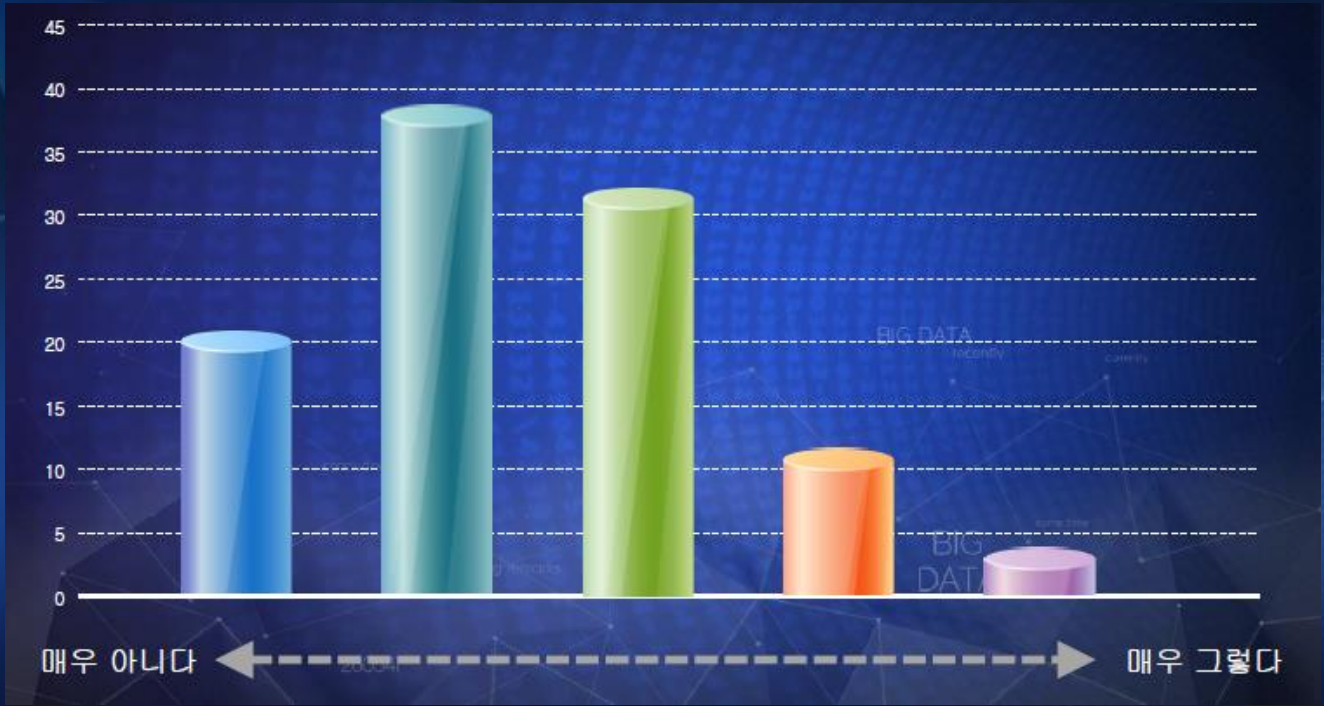
4. SK 텔레콤 설문조사 (4)

- 귀하의 '데이터 비식별화'에 대한 사전 인지도 수준은 어떠하신가요?



4. SK 텔레콤 설문조사 (5)

- 행정자치부의 '개인정보 비식별조치 가이드라인' 에 대해 들어본 적이 있으세요?

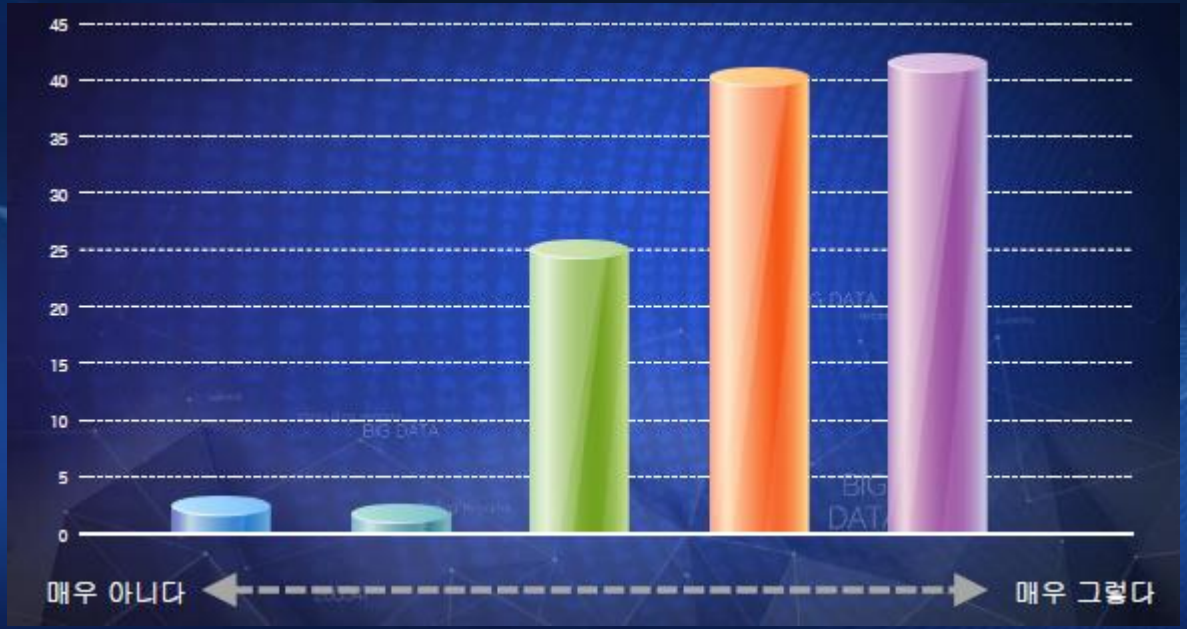


4. SK 텔레콤 설문조사 (6)

- 비식별 처리된 데이터가 귀사를 포함한 산업계 전반에 새로운 부가가치를 창출할 수 있다고 생각하시나요?

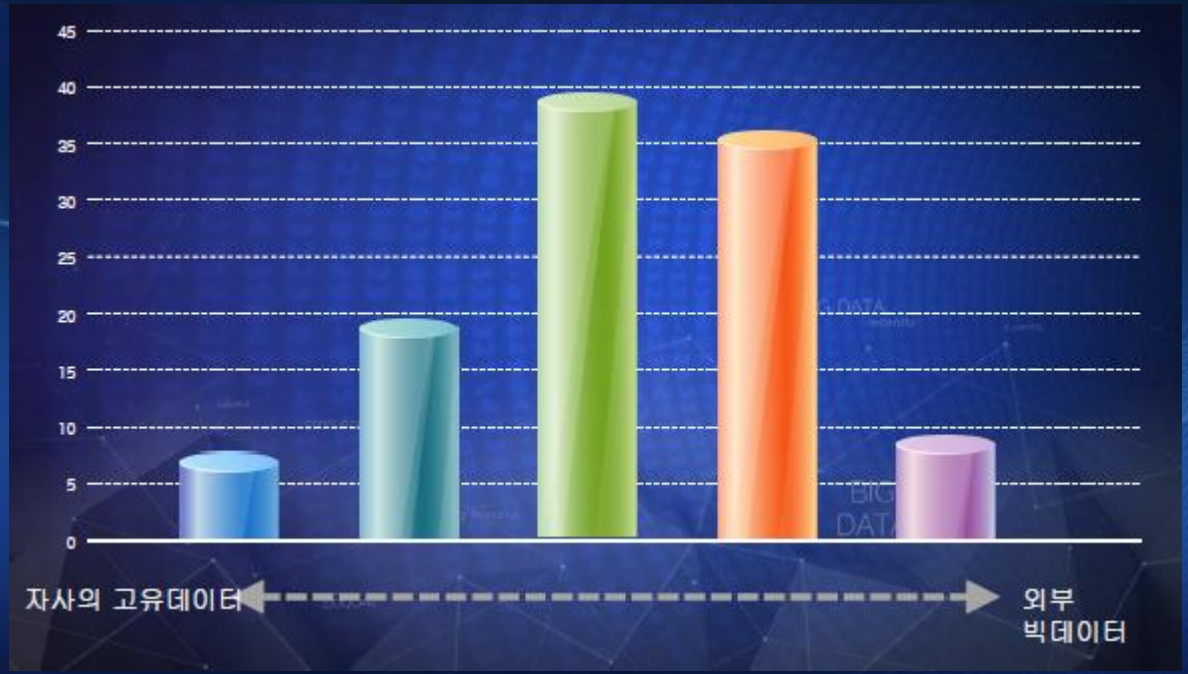
4. SK 텔레콤 설문조사 (7)

- 향후 귀사의 비지니스를 위해 데이터 활용의 중요성은 어떻게 고려되시나요?



4. SK 텔레콤 설문조사 (8)

- 귀사의 고유 데이터와 외부의 빅데이터 활용측면 중요도 비중은 어떻게 고려되시나요?

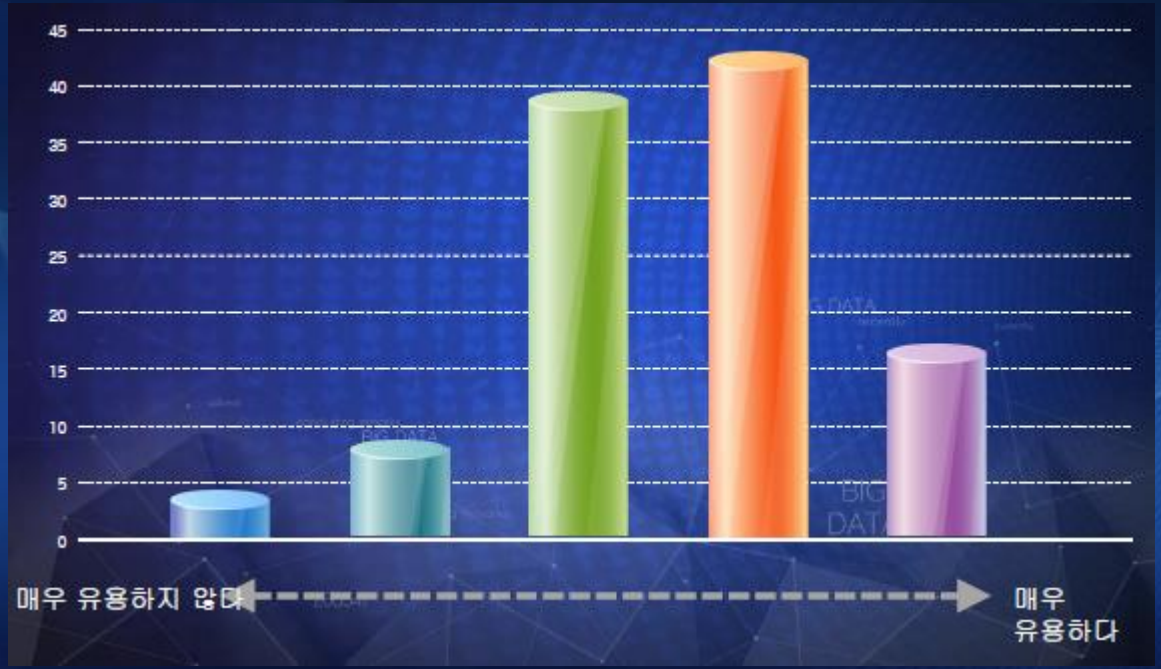


4. SK 텔레콤 설문조사 (9)

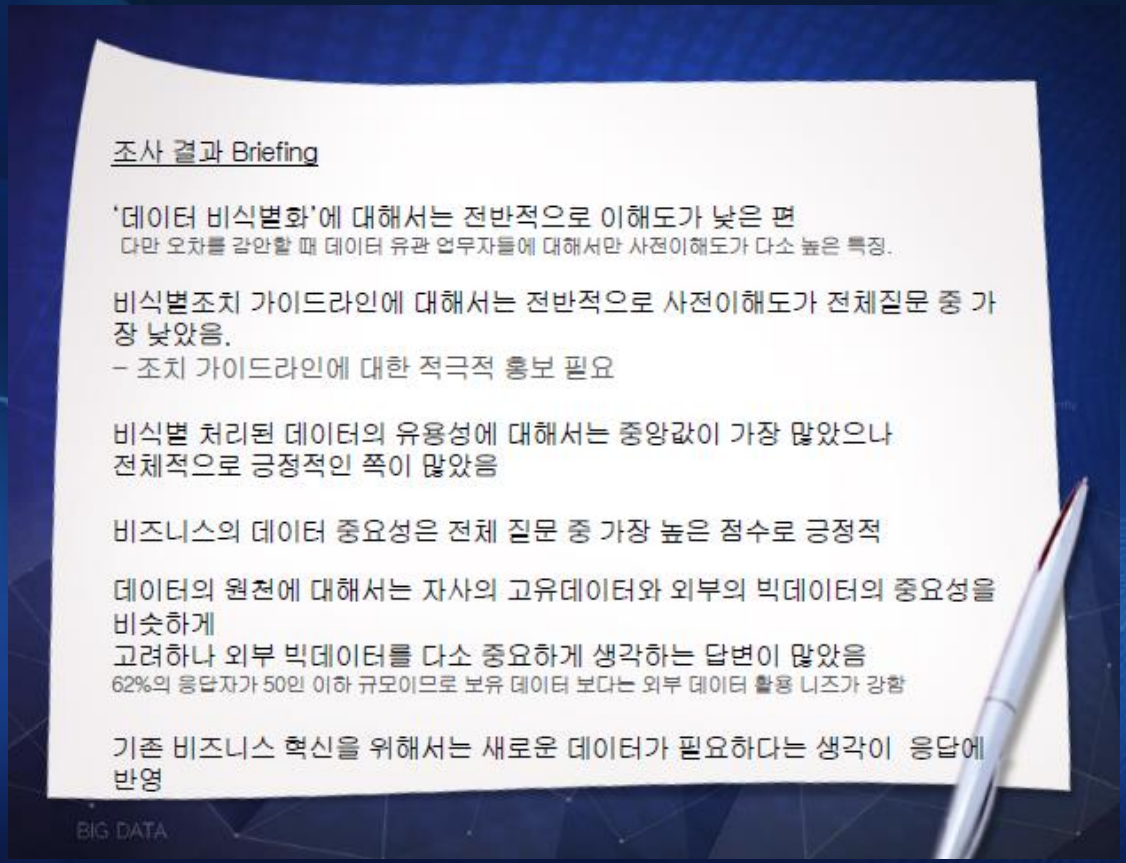
- SKT로부터 데이터를 서비스를 받을 수 있다면 어떤 내용에 관심이 있으세요? (중복선택 가능)

4. SK 텔레콤 설문조사 (10)

- 앞에 선택하신 데이터가 귀사의 비지니스에 얼마나 유용하다고 생각하시나요?



4. SK 텔레콤 설문조사 (11)



감 사 합 니 다

K-ICT 빅데이터센터