

CIO Summit 2019

Digital 2019, 연결하고 해석하고 지능화하라!

2019. 2. 21(목)

인터컨티넨탈 서울 코엑스 하모니볼룸(B1)



정보보안?

넘쳐나는 과·오탐을 막아야 한다.

서성춘 이사
세리정보기술(주)



- ① 변화의 물결, 디지털!
- ② 대응전략의 핵심은 '데이터분석'
- ③ 데이터분석 도구의 통합·연계
- ④ K사 보안모니터링 분석 적용사례



변화의 물결, 디지털!

디지털 쓰나미가 몰려옵니다.



고용

- » 단순·반복 업무의 자동화(일자리 양)
 - 힘들고 위험한 업무 자동화 및 양질의 일자리 증가
- » 고부가가치 업무로 재편(일자리 질)
 - 자동화가 어려운 창의·감성 업무로 노동의 가치 상승
- » 비전형 고용 확대 및 노동자의 근로선택 강화(고용형태)
 - 노동 시간·장소, 고용주에 종속되지 않는 대중노동(Cloud Work/Gig Economy) 확산

산업

- » 경쟁원천: 데이터
 - 기계의 자가학습에 필요한 데이터가 새로운 경쟁원천
 - 대규모 데이터를 확보하는 글로벌 ICT 기업이 시장 주도
- » 경쟁방식: 플랫폼 생태계
 - ICT 플랫폼과 연결된 다양한 서비스/제품으로 시장확장, 산업간 연결, 이종산업 침투
 - 더 많은 사용자를 확보한 대규모 플랫폼 기업이 경쟁우위 -> 승자독식

사회

- » 삶의 편의성 향상(헬스케어, 자율자동차, 로봇)
- » 안전한 생활 환경(보안·안전, 재난·국방)
- » 맞춤형 서비스(고객경험)
- » 해킹·양극화·개인정보 유출 등 역기능

“보안 분야에서의 디지털은 융합이며, 데이터 보안이 핵심”

오늘날 널리 보급되는 '디지털'은 보안분야에서 융합을 의미한다. 데이터 보안이 핵심(기밀/무결/가용성)

Digital Security for the Pervasive Digital Presence



- » 정보 보안의 범위 확장(IT보안, 관리 보안, 물리 보안)
- » Data Science의 발달에 따라 위협의 통합 관리 가능
- » IT, ICS, IoT의 융복합으로 인해 통합서비스 플랫폼 출현
- » 가장 복잡한 사이버 보안의 Risk 관리 기반 활용
- » 기존 보안산업에 임베디드 보안, 융합보안 자동화 등 새로운 기술 요구
- » 데이터 분석이 핵심



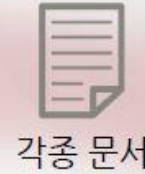
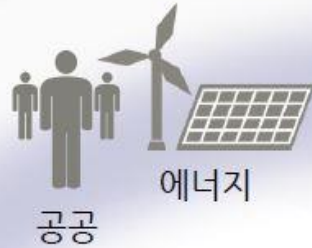
*출처: "Digital Transformation 시대의 사이버보안 전략" by SK인포섹 강용석 사업개발본부장



대응전략의 핵심은 '데이터분석'

빅 데이터
(비정형 데이터)

기간데이터
(정형 데이터)



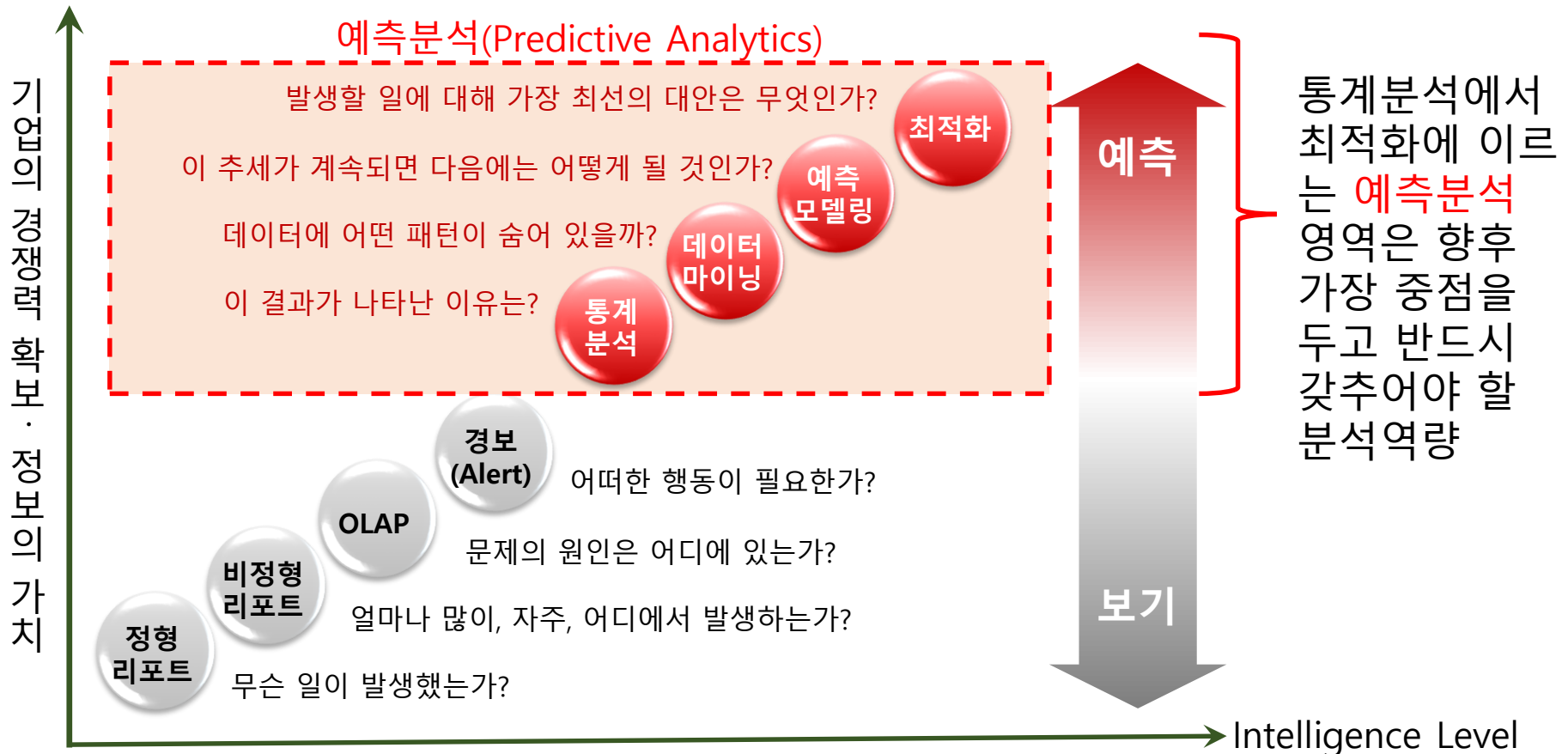
‘필수 다양성의 법칙(Law of Requisite Variety)’

“어떤 시스템을 완벽하게 통제하려면 통제 시스템의 다양성이 적어도 통제 받는 시스템보다 커야 한다.” 간단히 말해 ‘더욱 큰 다양성’만이 ‘다양성’을 통제할 수 있다는 이론.

“단순한 조회성 분석을 넘어 예측분석으로...”

정밀한 의사결정을 위해 경험으로부터 학습하여 미래행위를 예측한다.

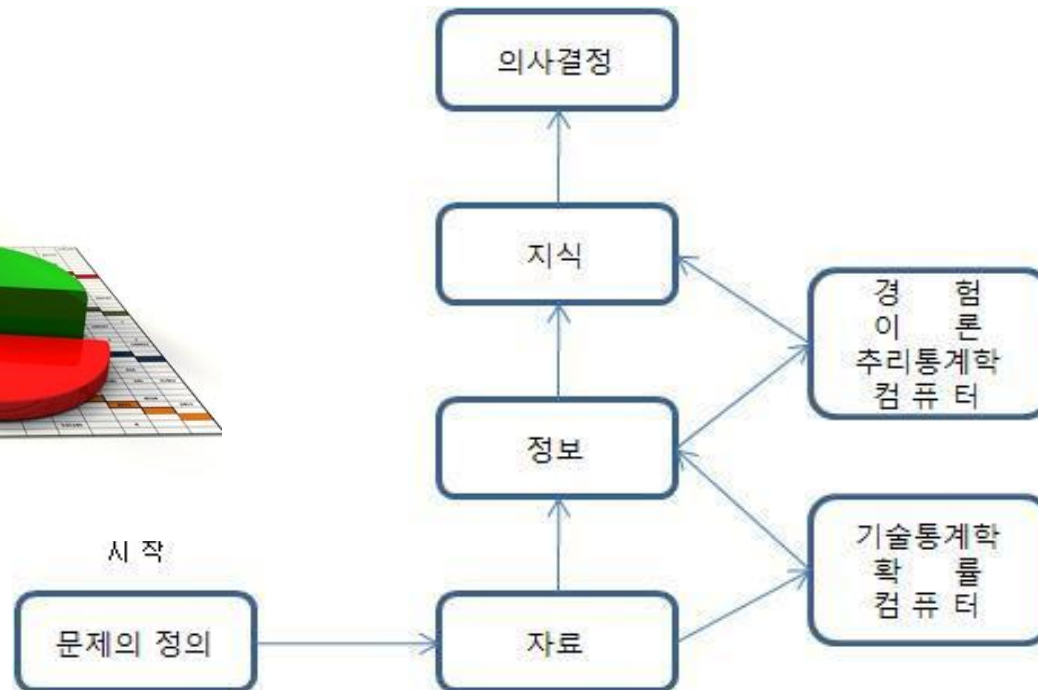
- Analytics 분야에서도 가장 고난이도 분야로, 단순히 데이터를 집계하고 시각화하는 분석의 수준을 뛰어넘어 **특정한 Event가 발생된 원인을 규명하고 향후에 발생할 가능성을 예측하며 이에 적합한 대응행동을 제시**해 줌으로써 데이터 기반의 합리적이고 효율적인 의사결정을 가능하게 해주는 미래형 분석 분야



“통계(Statistics) : 정형 데이터 분석”

자료를 수집하고 정리하여 이로부터 불확실한 사실에 대한 결론이나 일반적인 규칙성을 이끌어 내는 방법

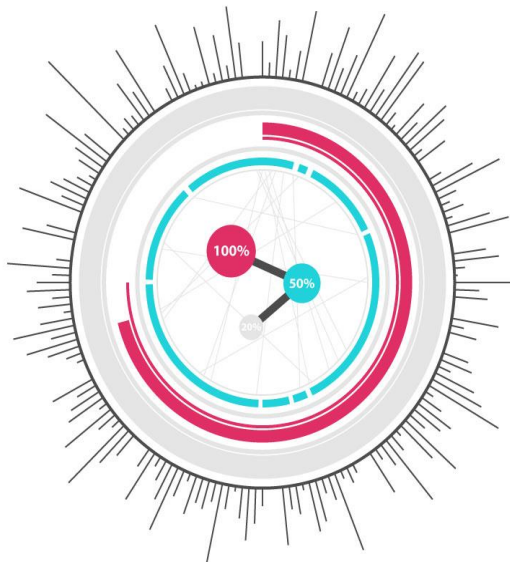
- 통계는 정형화된 대상 집단에 대한 모델링을 통해 일반성을 찾아내고 이것을 근거로 불확실한 사실에 대한 결론을 예측
- 자료를 수집/정리/분석할 뿐만 아니라, 그 분석을 토대로 불확실한 상황에서 현명하고 합리적인 의사결정을 할 수 있도록 하는 과학적 방법체계



“시각화(Data Visualization)”

데이터 분석 결과를 시각적인 표현과 도표라는 수단을 통해 명확하고 효과적으로 전달

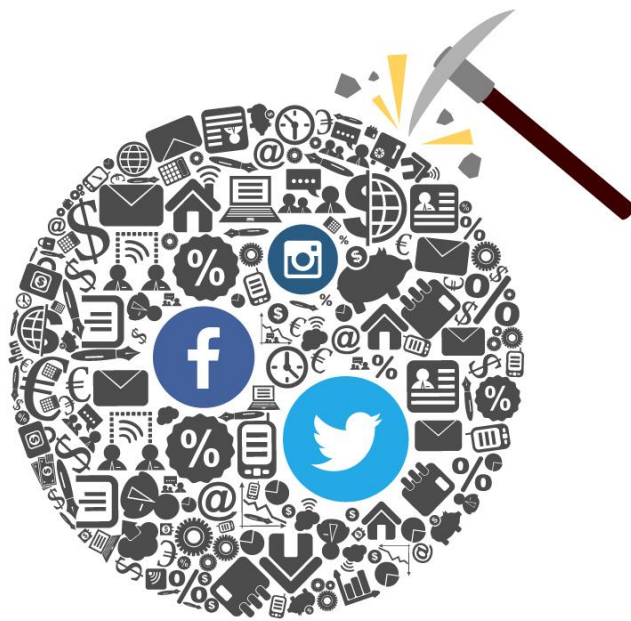
- 대규모 데이터에서 복잡함을 줄여주고 패턴과 상관관계를 탐색 가능하게 하며, 의미를 추론할 수 있게 하는 유용한 도구
- 사람의 감각 중 가장 많은 정보를 받아들이고 처리하는 '시각'을 중심으로 효과적인 정보전달을 위해 색, 그래프, 사진, 그림, 3D기술 등을 활용하여 데이터를 차별적으로 전달하는 작업
- “데이터시각화는 통계 분석 기법으로는 도저히 알 수 없는 데이터의 이야기를 끌어낼 것이다”
- 비주얼아이즈 디스



“텍스트마이닝(Text Mining)”

비/반정형 텍스트 데이터에서 새롭고 유용한 정보를 찾아내는 과정 또는 기술

- 비정형 텍스트 데이터에서 가치 있는 정보를 찾아내고, 다른 정보와의 연계성을 파악하며, 텍스트가 가진 카테고리를 찾아내는 등 단순한 정보검색 이상의 결과를 얻어낼 수 있음



(비정형)80 : 20(정형)
기업데이터 구성

주요기술

문서 분류
(Classification)

문서 군집
(Clustering)

정보 추출
(Information Extraction)

문서 요약
(Summarization)

활용분야

리스크 관리
(Risk management)

지식 경영
(Knowledge management)

사이버 범죄 예방
(Cybercrime prevention)

고객 관리 서비스
(Customer care service)

클레임분석을 통한 부정행위 탐지
(Fraud detection through claims investigation)

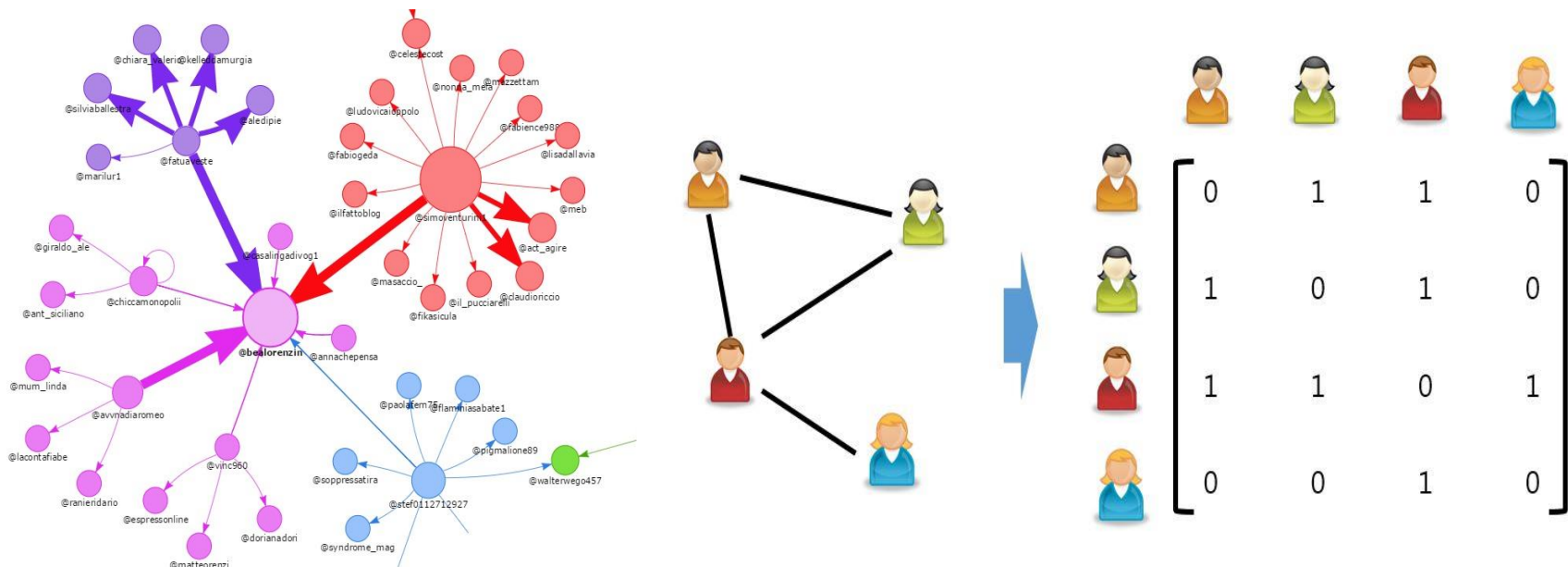
콘텐츠 강화
(Contents enrichment)

소셜 미디어 데이터 분석
(Social media data analysis)

“사회연결망분석(Social Network Analysis)”

‘관계(사람과 사람, 정보와 정보)’와 ‘상호작용’을 계량적으로 분석하여 미시적·거시적 관계 패턴을 파악

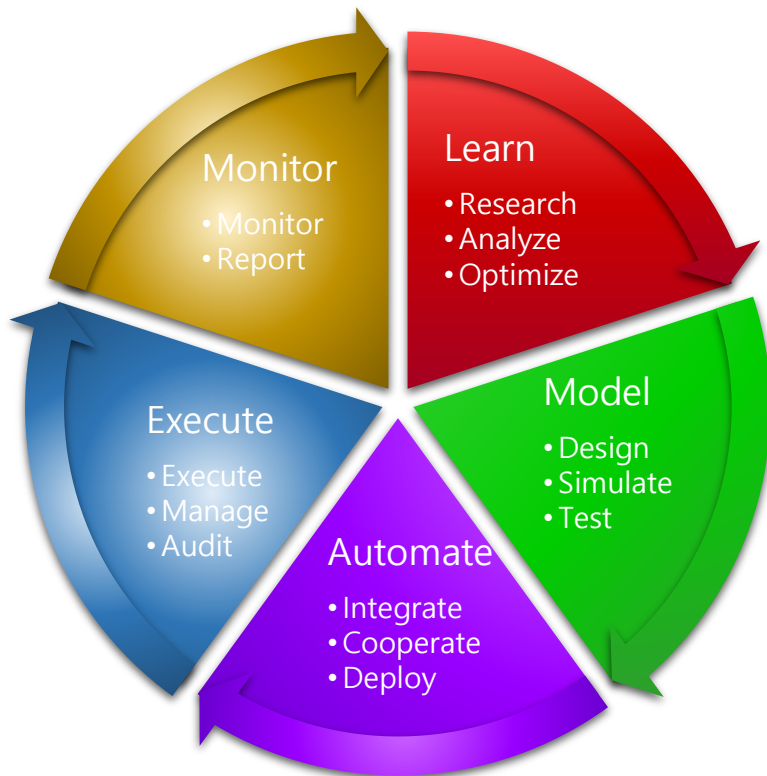
- 우리 주변에 존재하는 많은 것들은 네트워크 형태로 구조화되어 있거나 구조화할 수 있음
- 이에 대한 분석은 구성요소 간 상호의존성을 이해하고 전체의 효율성과 효과성을 증진하는 중요한 해결책을 제시
- 네트워크 내에서 차지하는 위치는 앞으로 마주치게 될 기회와 제약을 결정하고, 따라서 그 위치를 알아내는 것은 네트워크 구성원의 성과나 행태를 예측하는데 중요한 역할을 함
- 관계패턴에 초점을 맞춘 SNA는 전통적인 통계데이터 분석과는 다른 방법 및 분석적 개념을 요구(지위 접근법, 이벤트 접근법, 관계 접근법 등)



“BRMS(Business Rule Management System) : 비정형 데이터 분석”

복잡하고 다양한 결정 논리를 정의하고, 수행하며, 모니터링하고, 유지보수 하는 시스템

- 인간의 사고방식을 모방한 추론기능을 제공하여 경험과 지식이 반영된 시스템을 통해 지능적인 비즈니스 로직을 처리하여 의사결정을 지원



BR(Business Rule)

- 업무(Business)를 수행함에 있어서 필요로 하는 각종 규정, 원칙 등의 규칙
- 업무처리 지식과 Know-How를 포함
- 업무처리에 사용되는 모든 Rule을 이룸

BRE(Business Rule Engine)

- Business Rule을 자동으로 검색하여 처리하도록 하는 소프트웨어 Engine 및 Engine을 탑재한 시스템
- 복잡한 프로그램 표현 없이도 기업의 주요 Business Rule을 처리할 수 있도록 함

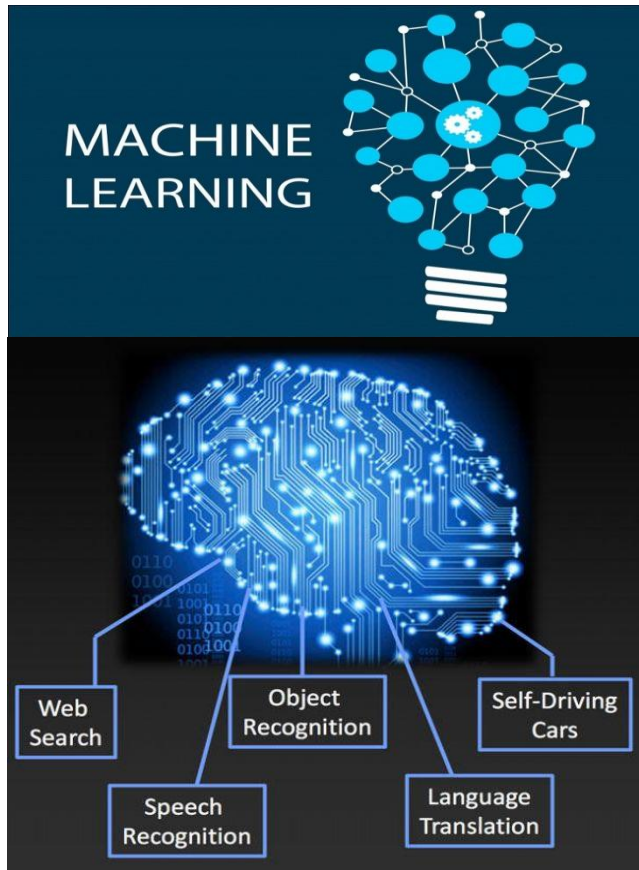
BRMS(Business Rule Management System)

- BRE를 포괄하는 개념
- 추론 기능을 보유한 BRE
- 기업의 핵심 Business Rule의 시스템화 및 활용을 개발자, 사용자, 운용/관리자의 3자 측면에서 지원하는 종합 Business Rule 관리환경

“기계학습(Machine Learning)”

인공지능의 연구 분야 중 하나로, 인간의 학습 능력과 같은 기능을 컴퓨터에서 실현하고자 하는 기술 및 기법

- 컴퓨터에게 사람이 직접 명시적으로 로직(Logic)을 지시하지 않아도 경험적 데이터를 통해 컴퓨터가 '학습'을 하여 패턴을 찾고 가장 효율적인 결과를 낼 수 있게 함



- **지도(supervised) 학습** : 미리 이용자가 만든 데이터를 입력한 뒤 출력까지 관여(데이터량이 많을수록 신뢰성 높음)
- **자율(unsupervised) 학습** : 출력 없이 입력만으로 패턴을 모델링(컴퓨터가 스스로 학습한 뒤 응용하여 출력, 대부분의 데이터 마이닝 기법에서 사용)
- **강화 학습(deep learning)** : 방대한 양의 데이터를 컴퓨터 스스로 학습한 후 생기는 피드백을 다시 학습하여 알고리즘을 생성(최적의 Action을 학습, 알파고)

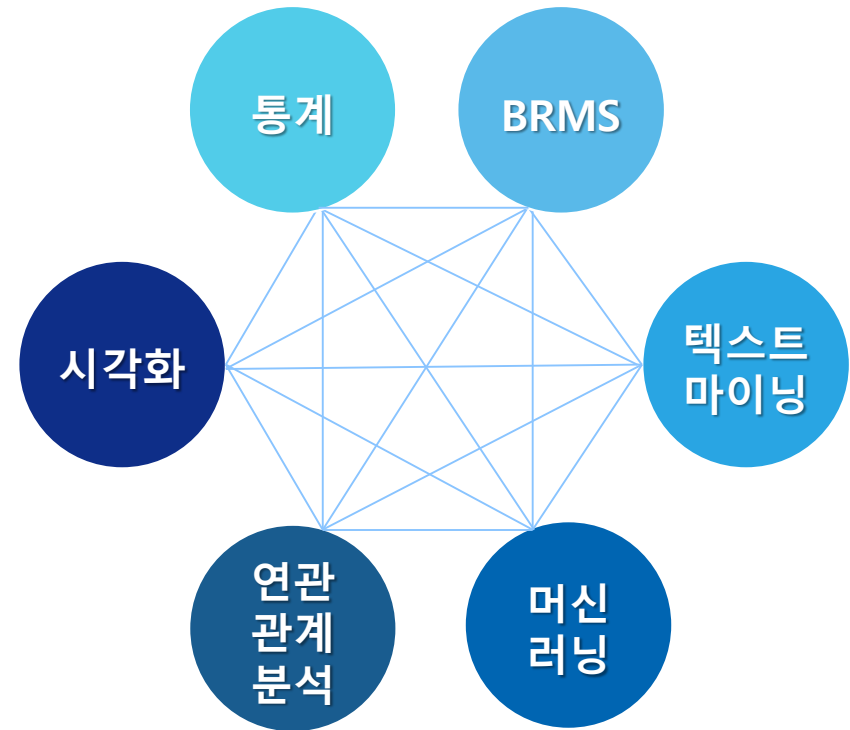
“어떠한 테스트(T)에 대해 꾸준한 경험(E)을 통하여 그 T에 대한 성능(P)을 높이는 것” - Tom M. Mitchell

기계학습에서 중요한 것은 E에 해당하는 데이터. 좋은 품질의 데이터를 많이 갖고 있으면 보다 높은 성능을 낼 수 있음.

“분석의 가치를 높이는 통합.연계”

개별 분석도구들이 통합되어 하나의 결과가 다른 분석도구와 연계되어 심화 분석 진행

- 고객의 신용상태를 확인하고 등급을 결정하여 차별화된 대출을 진행 (**통계 + BRMS + 시각화**)
- 특정 기간 동안의 제품구매 패턴과 소셜미디어 상의 제품 평판 정보를 연계하여 고객 마케팅 실시 (**통계 + 텍스트마이닝 + BRMS + 시각화**)
- 보험사기가 예상되는 지급 건에 대해 관련자(가해자/피해자/조사자/경찰/모집인/... 등)의 관계분석을 통해 보험사기여부 판단 (**연관 관계 분석+ BRMS + 통계 + 시각화**)
- 언론에 언급된 자사의 부정적 이슈에 대한 주가 변동을 예측하여 조기 대응 (**텍스트마이닝 + 통계 + BRMS + 시각화**)
- 음성/문자에 대해 고객과 상호작용을 통해 고객을 응대하고 지원 (**텍스트마이닝 + 머신러닝**)



“예측분석 및 의사결정지원 시스템”

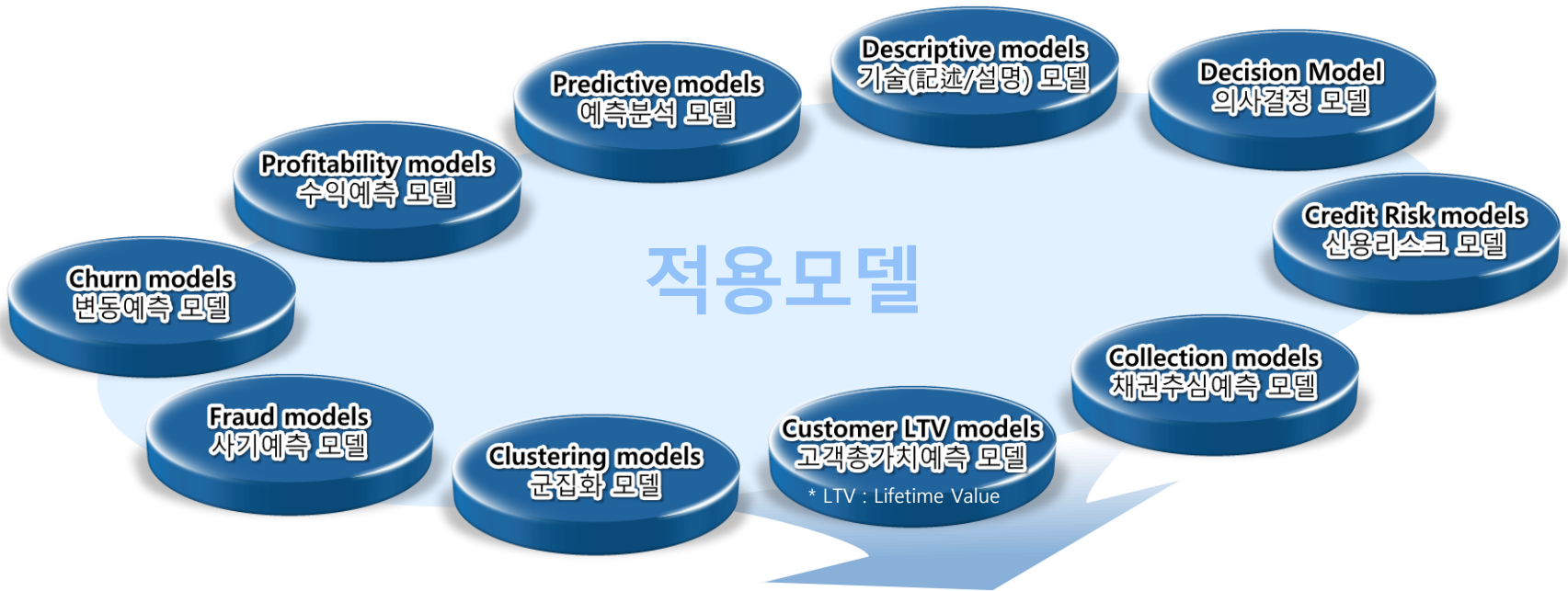
단일 아키텍처 내에 필요한 솔루션을 조합하여 운영, 확장성과 이식성이 높고 비용을 크게 절감할 수 있음



sFACT Framework

“DDD(Domain Driven Design) 방식의 재사용성”

견고한 모델이 구축되면, 다양한 산업분야의 다양한 업무에 동일한 모델을 적용할 수 있음



금융	보험	통신	유통	의료/보건	공공	제조	보안(공통)
<ul style="list-style-type: none"> 마케팅 고객관리 사기위험평가 신용위험평가 여신전문심사 채권회수/관리 	<ul style="list-style-type: none"> 계약전문심사 청구/지급심사 보험사기평가 여신전문심사 채권회수관리 	<ul style="list-style-type: none"> 마케팅 고객관리 거래사기 위험평가 대리점 관리 	<ul style="list-style-type: none"> 마케팅 위험평가 거래사기 지점관리 물류 최적화 	<ul style="list-style-type: none"> 청구전문심사 사기위험평가 이상징후평가 질병진단 자동화 	<ul style="list-style-type: none"> 구매조달 관세포탈 세금포탈 해외도피 사회서비스 평가 전장정보 지능화 	<ul style="list-style-type: none"> 구매조달 수요예측 물류 최적화 Optimization 	<ul style="list-style-type: none"> 개인정보보호 정보유출분석 위험행위분석 융복합위험분석 컴플라이언스

“기업 경영분석의 통합된 View 제공”

필요에 따라 제각각 도입/구축된 기존의 개별시스템간 통합.연계 분석의 어려움 해소

임원/ 경영자

- 기업 경영현황에 대한 **종합적 판단**
- 적시에 데이터에 기반한 신속하고 **정확한 의사결정**
- 경영정보 분석 역량 강화

현업 사용자

- 수작업 축소로 업무 효율 향상
- 다양한 관점의 분석업무 구현
- IT지원 없이 Self-serviced Discovery 실현

IT 담당자

- 데이터의 정합성, 신뢰도 향상
- 검증된 고성능의 분석 시스템 확보
- 일관되고 체계적인 관리 기반 구축
- 효과적인 권한관리 체계 마련

1 미래에 대한 통찰 제공

- 예측분석은 고객의 미래 행위에 대한 통찰을 제공
- 고객 또는 거래에 대한 **최적의 대응방법 제시**

2 복잡한 의사결정에 대한 과학적 수단 제공

- 기존: 복잡성이 높아 추측하기 어렵거나 어림짐작으로 처리
- 이후: **과학적 분석**에 의해 **신속한 처리**를 지원

3 불확실성 및 위험에 의한 비용 감소

- 분석적 통찰로 비즈니스 상의 위험을 정확히 측정
- 사기 및 비정상행위에 의한 **손실 감소**

4 컴플라이언스 및 대고객 서비스 효율 증진

- 개별 담당자의 주관적 편견 최소화
- 위험 통제 및 처리 지속성 등 **의사결정의 안정성 확보**

5 시장 경쟁력 확보에 기여

- 의사결정의 신속성, 지능화, 안정성으로 **빠른 시장 대응**
- 신규 시장에서의 대고객 서비스의 질 향상

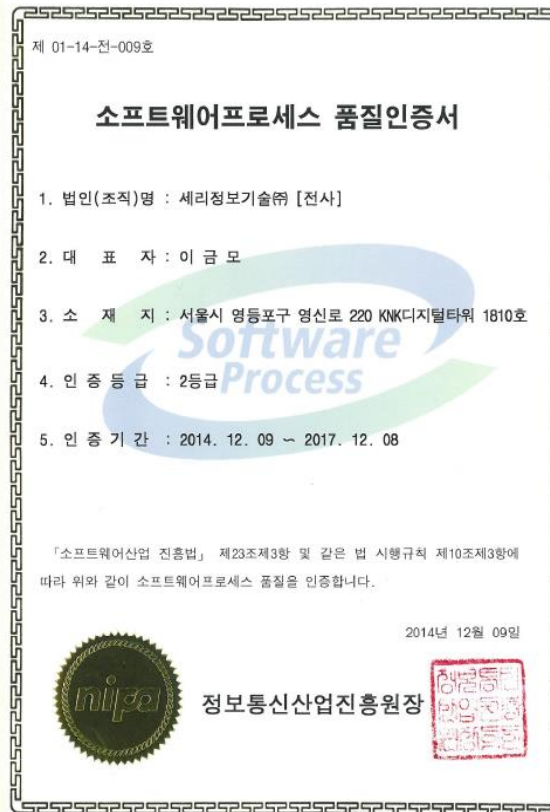
6 양질의 의사결정 선순환

- 작은 예측은 커다란 효과를 가져 옴
- 다양한 성공들이 모여 **지속적 성장**을 위한 선순환 제공

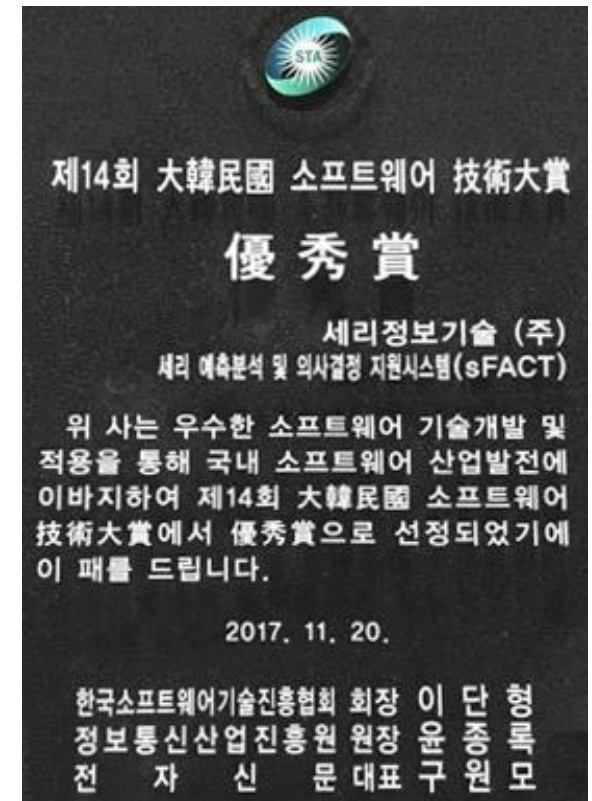
“GS 1등급, SP 2등급, 기술대상 수상” 공인된 기관으로부터의 기술력 인증



GS 1등급



SP 2등급



기술대상

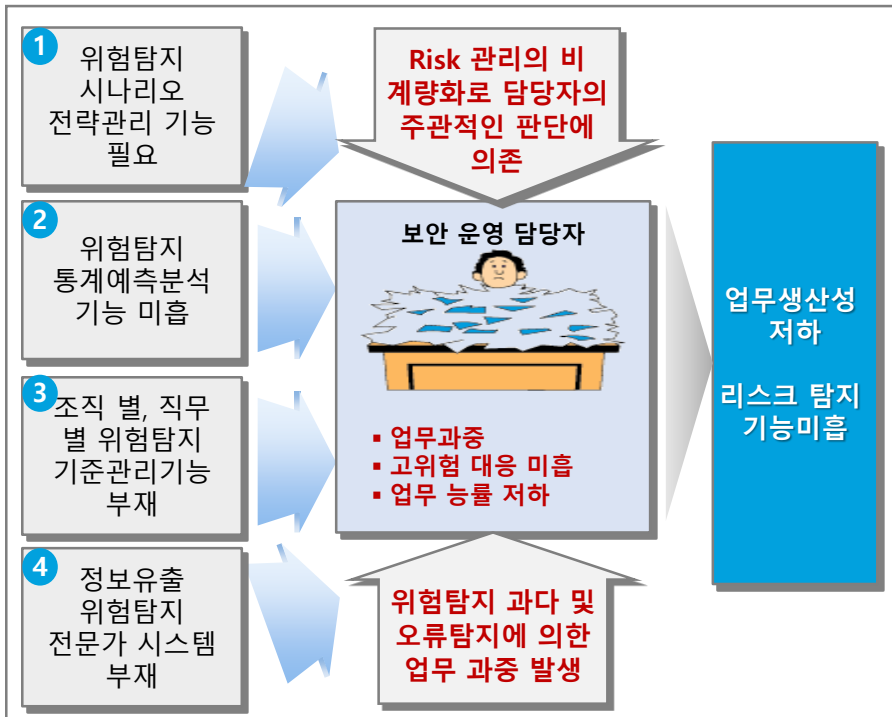


K사 보안모니터링 분석 적용사례

“위험탐지 과다와 오류탐지로 인한 업무과중”

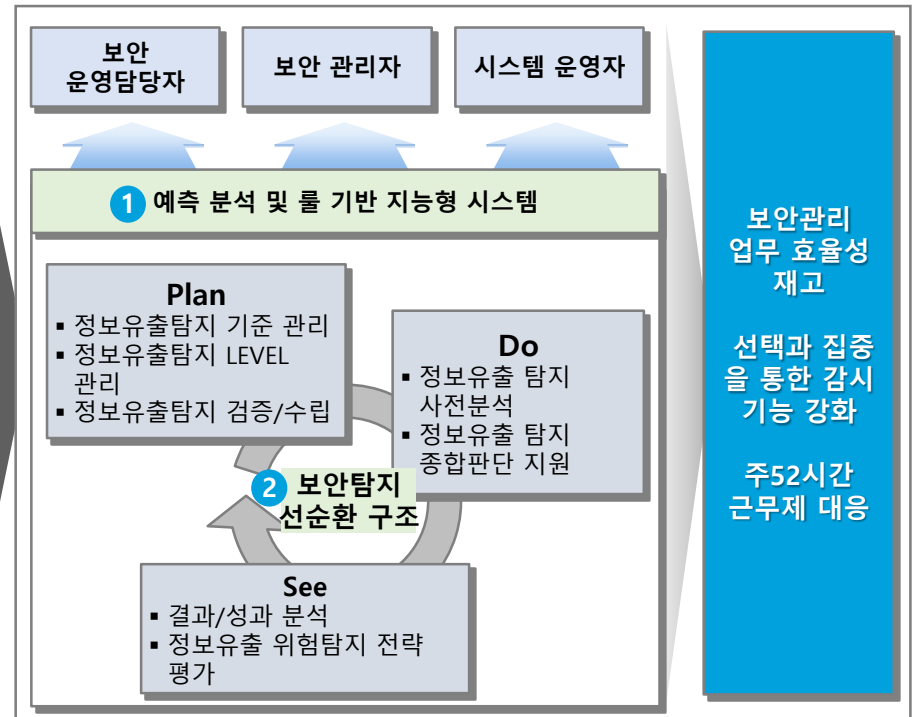
위험탐지기능 정교화와 정보유출탐지 종합판단 지원으로 업무생산성향상 및 선택과 집중을 통한 감시기능 강화

As-Is (현행 문제점)



- 1 정보유출 위험탐지 위험탐지 물량조절 기능이 없어 위험탐지 물량 과다와 오류탐지로 인한 업무 과중, 고 위험에 대한 대응 미흡
- 2 정보유출 위험탐지 판단 및 통계예측 분석을 위한 관리기능이 필요
- 3 조직, 직무 별 Top Secret 민감도 및 보안문서 History 관리 필요
- 4 보안관리업무 프로세스에 적합한 전문가 시스템 미흡

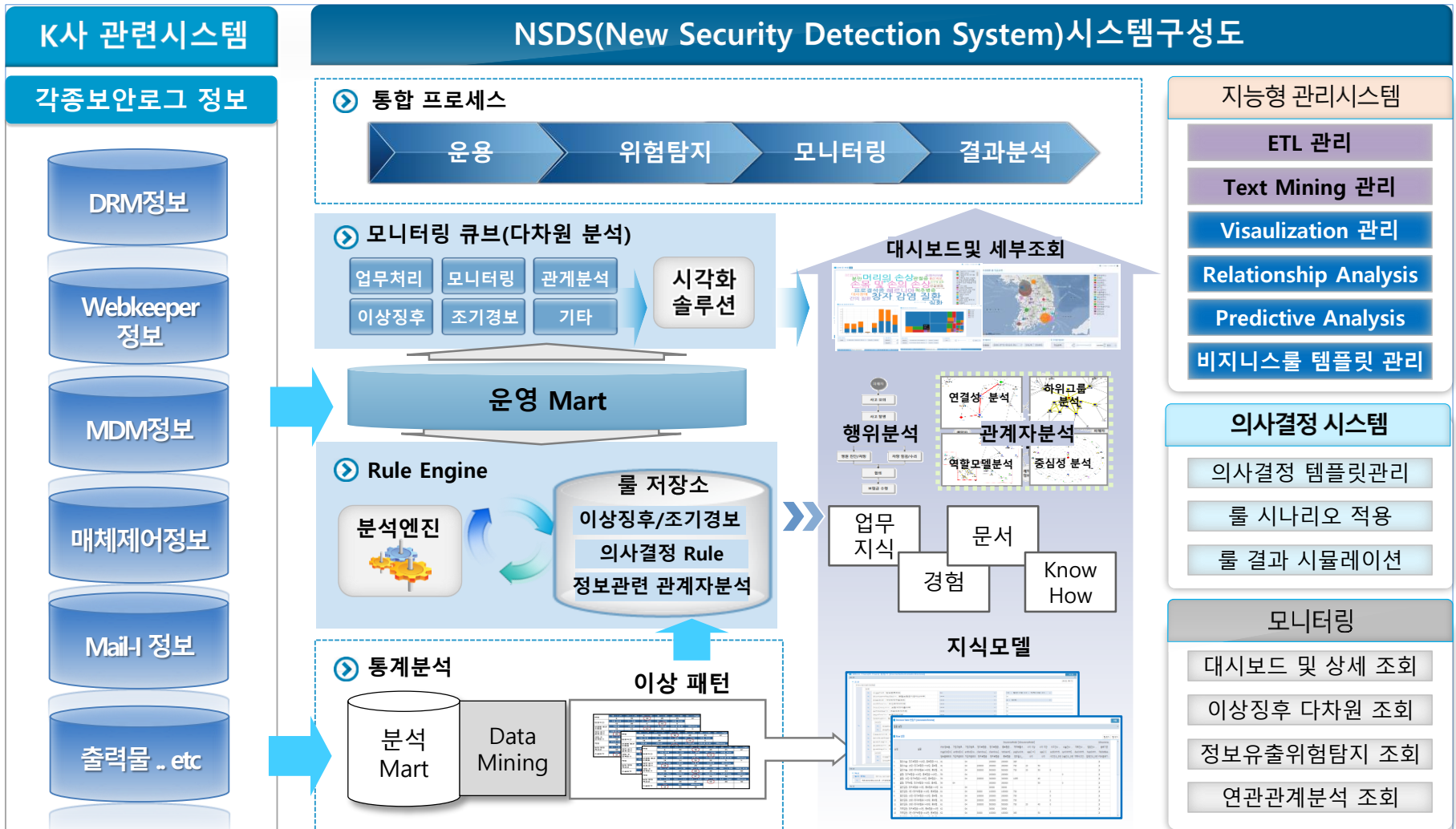
To-Be (개선사항)



- 1 통계모델에 의한 분석과 지식처리가 가능한 룰 기반 지능형 시스템을 적용하여, 보안 위험탐지 프로세스와 담당자에 맞는 업무 지원 기능 제공
- 2 보안관리업무의 선 순환 구조를 시스템화하여 체계적인 업무 수행을 지원하고, 단위 별 업무지식을 처리하여 자산화함으로써, 지속적으로 고도화/지능화 할 수 있는 기반 제공

"sFACT를 활용한 예측분석 및 의사결정지원 시스템 구성도"

단일 아키텍처 내에 필요한 솔루션을 조합하여 운영, 확장성과 이식성이 높고 비용을 크게 절감할 수 있음



“BRMS(Business Rule Management System) : 비정형 시나리오 관리” 템플릿을 활용한 다양한 시나리오 관리기능 제공

▶ 템플릿 목록	
템플릿 종류	템플릿 명
DecisionTable	DecisionTable 출력페이지
DecisionTable	DecisionTable 비밀문서
DecisionTable	DecisionTable 파일 삭제
DecisionTable	DecisionTable USB사용
DecisionTable	DecisionTable 암호화해제
DecisionTable	DecisionTable PC사용이력
DecisionTable	DecisionTable 출력횟수
DecisionTable	DecisionTable 암호화 해제 후 출력 리스크점수 추출
ScoreCard	ScoreCard 리스크점수 총합도출
DecisionTree	DecisionTree PC사용 리스크점수 추출
DecisionTree	DecisionTree 최종점수 분류

순번	설명	데이터모델		\$데이터모델
		파일 삭제 개수 ...	파일 삭제 개수 ...	파일 삭제 개수 ...
		FILE_DELETE...	FILE_DELETE...	FILE_DELETE...
1			2309	0
2		2309	2853	3
3		2853	3881	5
4		3881	4531	7
5		4531		10

CSV양식다운로드 | CSV업로드 | + 스코어 항목

간 출력 페이지 ... 주말 출력 페이지 ... 비밀문서 출력[데이... 파일 삭제 개수[데... 주간 USB 사용 ... 주말 USB 사용 ... 암호화 해...

스코어 항목

이름 암호화 해제신청 자가승인[데이

Fact(데이터 모델)
 데이터모델

필드명
 암호화 해제신청 자가승인[DOC_APP_ALLOW_SELF] : int

속성 속성 추가 | 속성 제거

연산자	값	구간 점수
<	2	
>=..<	2, 3	
>=	3	

1. DecisionTable(개별 위험 점수 산정)
 - 그리드(Grid) 형태의 룰을 사용하는 템플릿
2. ScoreCard(종합 위험 점수 산정)
 - 구간 및 항목들에 대한 점수를 측정하는 템플릿
3. DecisionTree(점수 구간에 대한 위험 등급 산정)
 - 트리(Tree) 형태의 룰을 사용하는 템플릿

“시각화(Data Visualization)”

정보유출 탐지에 필요한 각종 분석주제영역별 지표를 시각화를 통한 다차원분석 기법으로 정보 지원

시각화 관리메뉴

이상징후 분석주제 별 다차원분석
매체채널, 부서별, 출력, 암호해제, 파일발송, 복사 ...

- 해제신청
- 해제승인
- 매체
- 출력
- SNS발송
- 메일발송
- 웹하드
- USB
- ...

일자
근목의 시간
자가승인
권한자
Top Secret
....

검색조건 필터링

수치형 변수 통계함수처리

생성된 차트 대시보드 구성

대시보드 예시

위험리스크 점수 구간

위험점수 상위 직원그룹

위험점수 부서별 상위 그룹

일차별 USB 사용

10월 01일 USB 주간 사용자

10월 17일 USB 주간 사용자

주간 USB 사용 횟수

야간 USB 사용 횟수

업무의 PC사용이력

비밀문서 출력이력

주말 프린터 사용이력

아이디별 암호화해제 신청 및 자가승인

시각화 제공차트 기능

축 기반 차트

Non-Axis 차트

복합 차트 및 기초 통계함수 지원

지도 기반 Symbol 차트

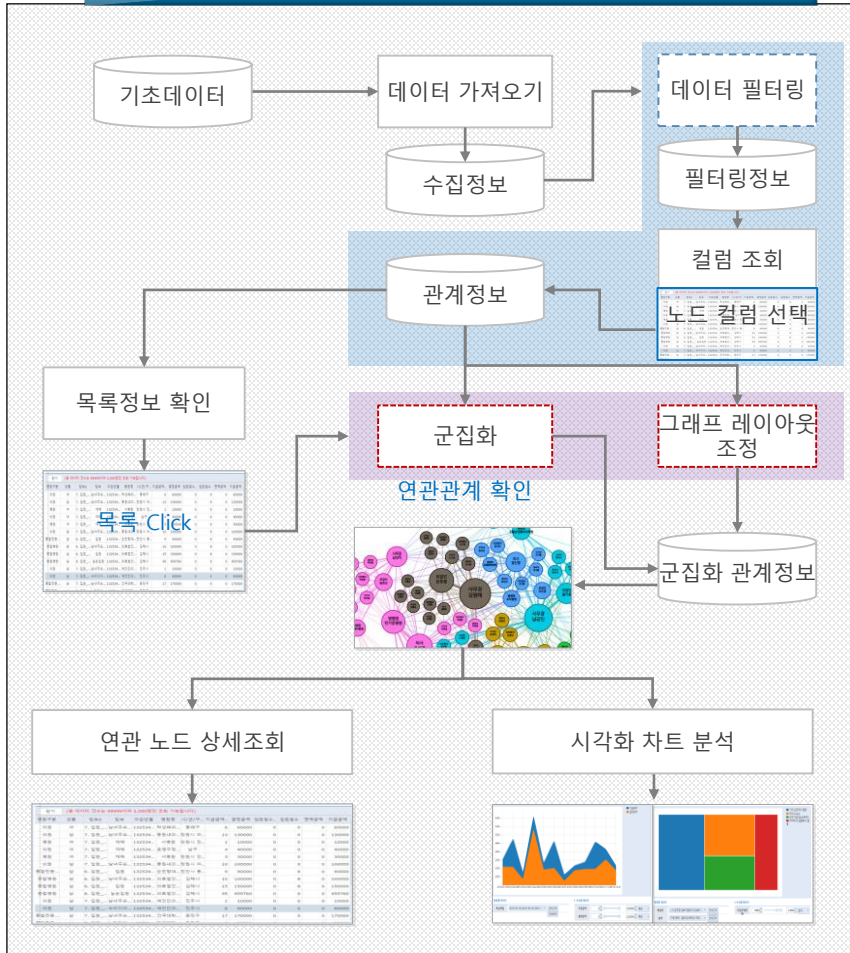
기초통계함수 지원

1. 평균
2. 중앙값
3. 분산
4. 표준편차
5. 최빈값

“사회연결망분석(Social Network Analysis)”

문서 암호화 신청 및 해제 직원간 연관관계 이상행위 패턴 분석으로 상시 모니터링 기능 강화

SNA(Social Network Analysis) 수행 절차



주요 내용

- ➔ sFACT 분석엔진의 SNA 솔루션 활용
- ✓ SNA개발 절차대로 템플릿을 제공함으로 개발생산성 및 유지보수 용이
 - ✓ 데이터의 필터링 기능을 활용하여 선택과 집중을 높임(관계 빈도수 높은 건)
 - ✓ 노드별로 선택하여 상세 조회기능을 제공함으로 분석이 용이함
 - ✓ 연관관계화면에서 노드별로 빈도/지지도/신뢰도 기능 제공

DRM 문서 암호화 해제 직원간 연관관계



도출된 비정상 패턴(예시)

- 연관관계분석
- 직원 간 관계 (요청자/승인자 구분)를 시각적으로 분석
- 높은 영향력을 가진 노드 (Node)는 큰 사이즈로 표시
- 드릴다운
- 개별 노드와 관련된 연관데이터 확인 가능

“텍스트마이닝(Text Mining)”

Web접속이력을 통한 보안모니터링 분석

2018년 10월 사이트 접속목록



명목형 데이터

수치형 데이터

접속사이트명: 11번가, AK플, AeroMexico, AliExpress, BITTREX, BitMEX, Booking.com, CJ오... 선택 해제

“통계(Statistics) : 정형 데이터 분석”

관리도 통한 위험 임계치 도출 및 위험도 산출기준은 정규분포를 이용한 구성비를 기준으로 위험점수를 산출

구성비 %	z값	산출값
10	1.28	9
5	1.64	10
1	2.32	13
0.3	2.75	14

정규분포 구성비 (분포 넓이 값)

산출값 = 평균 + (z 값 * 표준편차)

02. 과다사용

출력횟수별 점수(일 기준 몇회)

회사총괄 (SAFER)

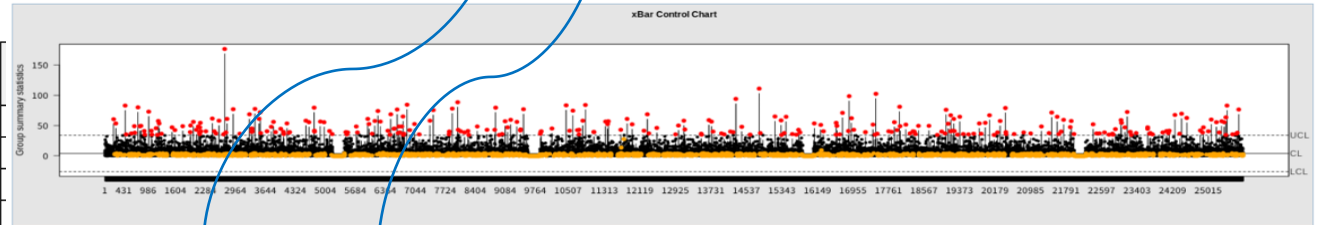
점수	이상	이하	가중치	가중치 환산	구성비 %	평균	표준편차
0		9	1	0	10 < ,	4.609	3.731
3	9	10	3	5 < , < 10			
5	10	13	11	1 < , < 5			
7	13	14	1	0.3 < , < 1			
10	14		10	< 0.3			

출력페이지별 점수(일 기준 페이지수)

회사총괄 (SAFER)

점수	이상	이하	가중치	가중치 환산	구성비 %	평균	표준편차
0		26	1	0	10 < ,	11.921	11.093
3	26	30	1	3	5 < , < 10		
5	30	37	1	5	1 < , < 5		
7	37	42	1	7	0.3 < , < 1		
10	42		1	10	< 0.3		

Xbar 관리도 그래프



통계치

통계 항목	평균	표준편차	최대값	최소값	UCL(xbar)	LCL(xbar)	UCL(sub)	LCL(sub)	데이터 길이
통계치	4.609	3.731	176.5	0.0	34.1697	-25.8501	-	-	25804

경고 및 위험

idx	위험군_index	위험군_값	경고군_index	경고군_값
1	197	60.5	6306	5.0
2	246	53.5	11702	4.5
3	458	83.0	11703	13.5
4	482	35.5	11764	27.5
5	598	37.0	14105	4.5
6	666	49.0	16241	9.0
7	678	38.0	17239	39.0
8	749	80.0	224	3.0
9	793	49.0	225	2.0
10	816	49.5	264	4.0

Table AN-2 Standard Normal Table

P(0 < Z < 1.55)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9340	0.9348	0.9356	0.9364	0.9371	0.9379	0.9386	0.9394	0.9401

정규분포 z값 표

“Optimization(최적화)”

모니터링, 분석, 리모델링, 검증, 테스트, 적용의 사이클을 통한 최적화 수행

정보유출 탐지모델 모니터링

- 비즈니스 룰 모니터링
- 모델 룰 모니터링
- 탐지 추이 분석
- 모델 성능 분석



정보유출 탐지모델 전략적 적용



모집단 검증 및 테스트



정보유출 탐지모델 분석

- 위험등급별 영향도 분석
- 프로세스별 영향도 분석
- 변수별 영향도 분석

Optimization(최적화)

시뮬레이션 및 검증

- 단위 데이터 시뮬레이션 및 검증
- 복수 모델 비교 시뮬레이션 및 검증



리모델링

- 비즈니스 룰 개발: 신규 패턴 추가 및 비즈니스 룰 생성
- 모델 룰 개발: 신규 통계 Rule 개발, 신규 시나리오 개발, 모델 룰 등록

최적화



모델결과 영향도 성능 분석

리모델링을 통한 시간단축

모델검증 시뮬레이션

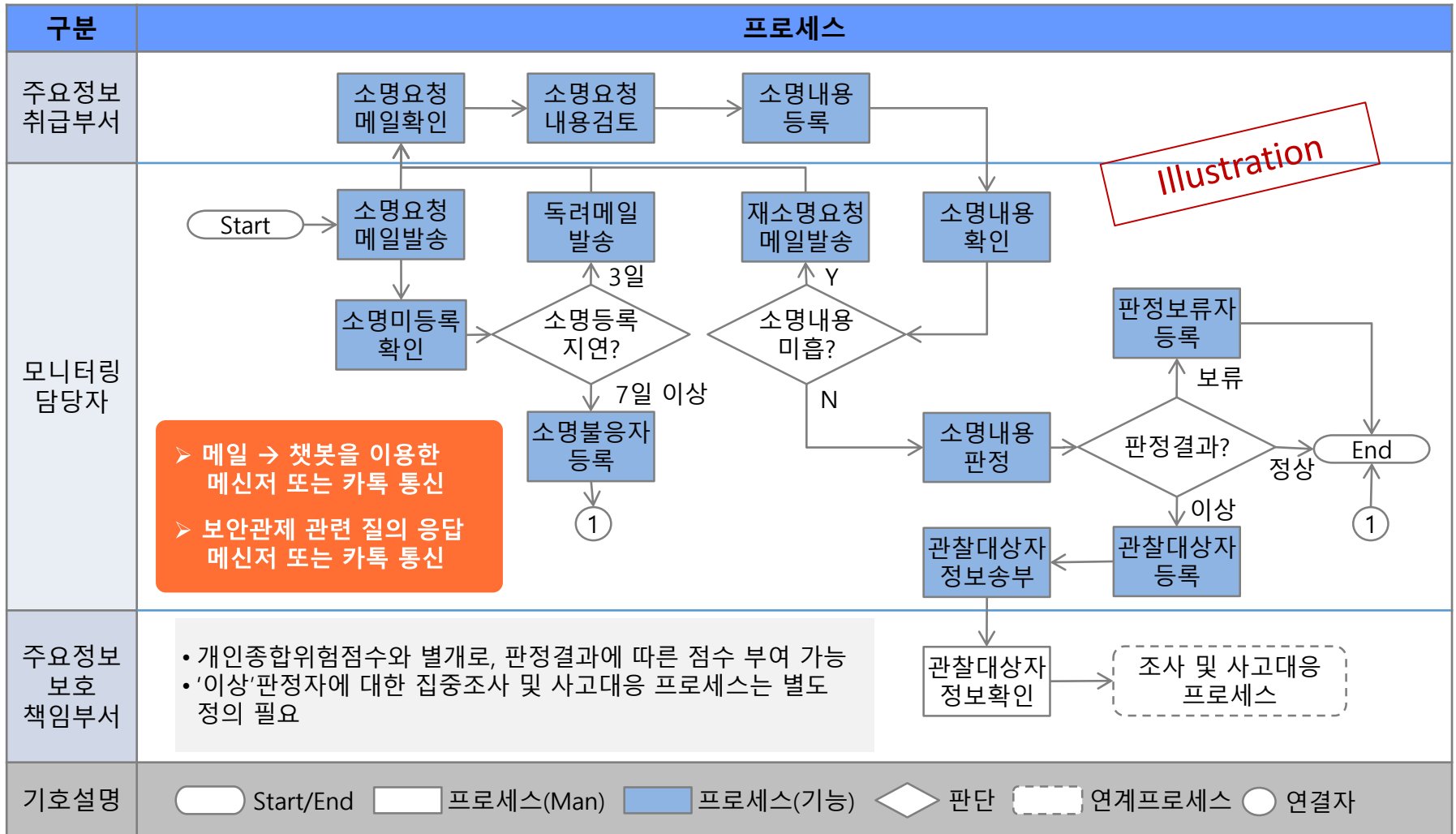
비교모델 검증을 통한 모델적용

신-구 전략모델 운영

모델 선 순환을 통한 최적의 운영기능개발

“챗봇 시스템 도입”

소명요청 및 보안관제 관련 관리규정 질의응답을 위한 커뮤니케이션 창구로 챗봇을 활용



예측분석을 위한 통합 플랫폼 : **sFACT™**

THANK YOU

Q&A



서 성 춘 이사

www.seritech.co.kr

서울시 영등포구 영신로220 KnK디지털타워 18층

scseo@seritech.co.kr