

IV. 빅데이터 결과 해석

02. 분석 결과 해석 및 활용

2.1 분석 결과 해석

-	KeyWord
분석 모형 해석	데이터시각화, 시간시각화, 분포시각화, 관계시각화, 비교시각화, 공간시각화, 시각화도구, 태블로, 인포그램, 차트블록, 데이터래퍼
비즈니스기여도평가	총소유비용(TCO), 투자대비효과(ROI), 순현재가치(NPV), 내부수익률(IRR), 투자회수기간(PP)

=====

1) 분석 모형 해석

(1) 데이터 시각화(Data Visualization)

- 데이터에 대한 이해를 돕기위해 그림/도형과 같은 그래픽 요소들을 이용하여 데이터를 묘사/표현하는 과정
- 데이터 시각화 방법
 - 선/막대/원 등의 기하나 도형과 같은 양식을 이용하여 데이터 특징을 설명할 수 있는 모양으로 만들거나 색상/레이블 등의 특성을 활용하여 데이터를 표현
- 데이터 시각화 기능 구분: 설명 / 탐색 / 표현 기능
 - 설명 기능: 전달하려는 메시지, 주요한 분석결과를 설명 -> 사용자에게 명확하게 보여줘야 함
 - 탐색 기능: 데이터에 숨겨진 관계와 패턴을 찾기 위함 -> 사용자가 직접 탐색함
 - 표현 기능: 이야기 전달과 공감을 불러일으키기 위함 -> 예술적 표현, 작품
- 데이터 시각화의 목적
 - 사용자가 주제에 대해 더 잘 이해하고 느끼게 하는 것
 - 정보 전달: 전달 & 분석 -> 실용적 & 과학적인 측면
 - 설득: 공감 & 설득 -> 추상적 & 예술적인 측면
- 데이터 시각화 절차: 구조화 -> 시각화 -> 시각표현
 - 구조화: 목표 설정/데이터 표현 규칙&패턴 탐색/시각화 요건정의/시나리오&스토리 구성
 - 시각화: 시각화 도구&기술 선택/시각화 구현
 - 시각표현: 그래프 보정&품질향상/전달요소 강조/인터랙션 기능 적용/결과물 검증

(2) 데이터 시각화 유형

- 시간 / 분포 / 관계 / 비교 / 공간 시각화

시간 시각화	분포 시각화	관계 시각화	비교 시각화	공간 시각화
시간 흐름에 따른 변화 경향/트렌드 파악	분류에 따른 변화를 최대/최소/전체분포 등으로 표현	집단간 상관관계 확인 → 다른수치 변화 예측	데이터간 차이점/유사성 관계확 인	지도를 통해 시점에 따른 경향/차이 확인
막대 그래프 점 그래프	파이 차트 도넛 차트 트리맵	산점도 버블차트 히스토그램	히트맵 평행 좌표 그래프 체르노프 페이스	등치선도 도트맵 카토그램

(3) 시각화 분석을 위한 데이터 유형

- 범주 및 비율 / 추세 및 패턴 / 관계 및 연결
- 범주 및 비율: 범위 / 분포 / 순위 / 측정
- 추세 및 패턴: 추세 방향 / 추세 패턴 / 추세 속도 / 변동 패턴 / 중요도 / 교차
- 관계 및 연결: 예외 / 상관성 / 연관성 / 계층 관계

데이터 유형: 범주 및 비율			
범위	분포	순위	측정
값의 범위 파악	개별 변수&변수 조합이 가지는 분포 형태 파악	크기 기준으로 순서 확인 최댓값/최솟값/ 중위수/사분위수	값이 가지는 중요성 파악 숫자 자체보다 깊이있는 조사!

데이터 유형: 추세 및 패턴					
추세 방향	추세 패턴	추세 속도	변동 패턴	중요도	교차
값이 증가/감소	선형/지수형 or 변화없음	어느 정도로 추세가 급한가	반복 패턴, 변동폭, 무작위 패턴 등	확인한 패턴이 중요 or 잡음 파악	변수간 교차(점), 중첩 발생 여부 확 인

데이터 유형: 관계 및 연결			
예외	상관성	연관성	계층 관계
이상값과 같은 정상범위 벗어난 변수 파악	변수간 관련성 강한/약한 상관관계 존재 확인	변수와 값의 조합간 의미있는 관계 파악	데이터 범주의 구성/분포/관련성 파악

(4) 빅데이터 시각화 도구

- 태블로 / 인포그램 / 차트블록 / 데이터래퍼
- 태블로(Tableau): 데이터를 클라우드에 저장 -> 차트/그래프/지도 등의 다양한 그래픽 기능 제공
- 인포그램(Infogram): 실시간으로 인포그래픽을 연동해줌
- 차트 블록(Chart Blocks): 코딩없이 데이터베이스 형태의 데이터를 쉽게 시각화 -> 웹기반 차트 구현
- 데이터 래퍼(Data Wrapper): 사용자 목적에 따라 제작할 수 있는 레이아웃 제공 -> 차트/맵으로 데이터를 변환

=====

2) 비즈니스 기여도 평가

(1) 비즈니스 기여도 평가

- 사업수행에 영향을 주는 요소를 수치화된 자료형태로 산출하는 평가방법

- 비즈니스 기여도 평가지표: 총소유비용/투자대비효과/순현재가치/내부수익률/투자회수기간

총 소유 비용	투자 대비 효과	순 현재가치	내부 수익률	투자 회수 기간
TCO	ROI	NPV	IRR	PP
한 자산을 획득하려 할 때, 주어진 기간동안 모든 연관 비용을 고려	자본 투자에 따른 순 효과의 비율 (투자 타당성)	특정시점의 투자금액과 매출금액의 차이를 이자율로 고려	순 현재가치를 0으로 만드는 할인율 (연단위 기대수익 규모)	누계 투자금액과 매출 금액의 합이 같아지는 기간 (흑자 전환 시점)

- 고려사항: 효과 / 성능 / 중복 / 최적화 검증
 - 효과 검증: 검출률 증가 / 향상도 개선 등의 효과 제시
 - 성능 검증: 시뮬레이션을 통한 처리량 / 대기시간 / 대기행렬 감소
 - 중복 검증: 타 모델링과의 중복에 따른 효과를 통제 / 제시할 수 있어야 함
 - 최적화 검증: 최적화를 통해 목적함수가 증가한 만큼의 효과 제시

2.2 분석 결과 시각화

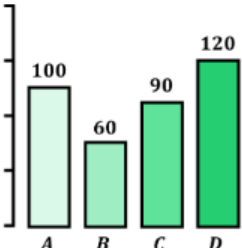
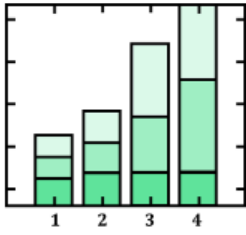
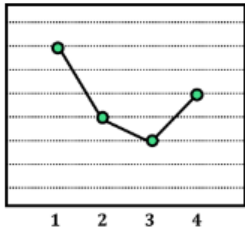
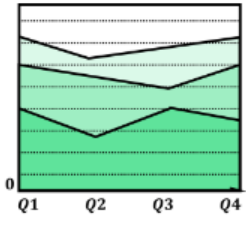
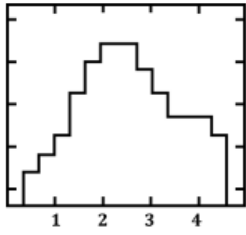
-	KeyWord
시공간 시각화	시간시각화, 막대그래프, 누적 막대그래프, 선그래프, 영역차트, 계단식그래프, 공간시각화, 등치 지역도, 등치선도, 도트맵, 도트플롯맵, 버블맵, 버블플롯맵, 카토그램
관계 시각화	산점도, 산점도 행렬, 버블차트, 히스토그램
비교 시각화	플로팅바차트, 히트맵, 체르노프페이스, 스타차트, 평행좌표그래프
인포그래픽	인포그래픽

=====

1) 시공간 시각화

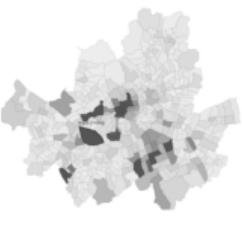
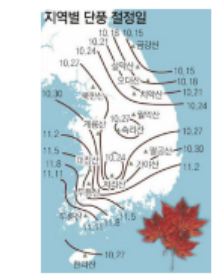

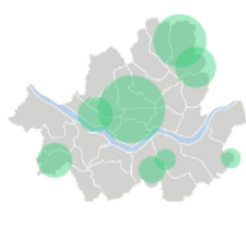
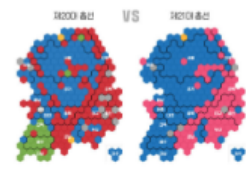
(1) 시간 시각화

- 시간에 따른 데이터 변화를 표현
- 주요 관심요소: 경향성/트렌드/추세
- 시간 시각화 유형: 막대그래프/누적 막대그래프/선그래프/영역차트/계단식그래프

막대그래프	누적 막대그래프	선 그래프	영역 차트	계단식 그래프
Bar Graph	Stacked Bar Graph	Line Graph	Area Chart	Step Line Graph
				
<ul style="list-style-type: none"> - 동일 너비 막대 여러개 - 막대: 특정 범주 - R: geom_bar() 	<ul style="list-style-type: none"> - 전체 비율 보여줌 - 여러 범주를 동시에 차트로 표현 가능 	<ul style="list-style-type: none"> - 점: 수량 - 점들을 선분으로 이음 - 시간에 따른 크기변화 	<ul style="list-style-type: none"> - 색을 채운 영역 - y축 값은 0부터 시작 - 시간에 따른 크기변화 	<ul style="list-style-type: none"> - 변화가 생길 때까지 x축과 평행한 선 유지 - 값이 변하는 지점에서 급격하게 뛰어오름

(2) 공간 시각화

- 지도 상에 해당하는 정보를 표현
- 대부분 위도&경도 사용함
- R에서 정적/동적인 방법으로 시각화 가능
 - 정적 공간 시각화 패키지: maps, mapproj, maptools, mapplots
 - 동적 공간 시각화 패키지: RgoogleMaps, ggmap
- 공간 시각화 유형: 등치지역도/등치선도/도트맵/도트플롯맵/버블맵/버블플롯맵/카토그램

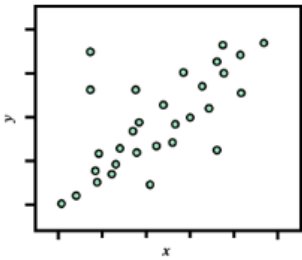
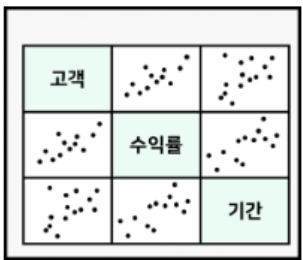
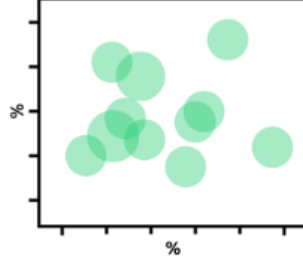
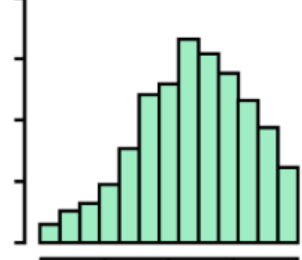
등치지역도	등치선도	도트맵/도트 플롯맵	버블맵/버블 플롯맵	카토그램
Choropleth Map	Isometric Map	Dot (Plot) Map	Bubble (Plot) Map	Catogram
				
<ul style="list-style-type: none"> - 색상으로 구분 - 데이터값 크기에 따라 채도/ 밝기 변화 - 인구밀도가 다른 경우 왜곡되는 결점 존재함.. 	<ul style="list-style-type: none"> - 등치지역도 결점 극복 - 데이터값 크기에 따라 색상 농도 변화 	<ul style="list-style-type: none"> - 산점도처럼 점으로 표현 - 활용: 시간 경과에 따라 점진적인 확산을 나타내는 경우 	<ul style="list-style-type: none"> - 데이터값 크기에 따라 서로 다른 크기의 원형으로 표시 	<ul style="list-style-type: none"> - 데이터값 크기에 따라 면적을 왜곡시킴 - 지리적 형상크기를 조절 - 재구성된 지도

=====

2) 관계 시각화

(1) 관계 시각화

- 다변량 데이터 사이에 존재하는 변수간 연관성/분포/패턴을 찾는 시각화 방법
- 상관관계를 시각화하는 기법
 - 한 가지 요소의 변화가 다른 요소의 변화와 관련이 있는지 표현하는 시각화 방법
- 관계 시각화 유형: 산점도/산점도 행렬/버블차트/히스토그램

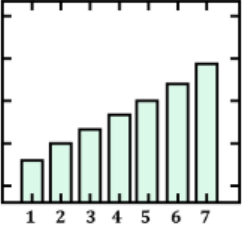


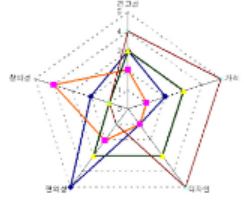

산점도	산점도 행렬	버블 차트	히스토그램
Scatter Plot	Scatter Plot Matrix	Bubble Chart	Histogram
			
- 두 변수 각각 X축, Y축 - (X, Y) 순서쌍이 하나의 점! - R: geom_point() - 상관관계/ 군집화/ 이상값 패턴 파악에 유용함	- 다변량 변수 데이터에서 - 가능한 모든 변수 쌍에 대한 - 산점도를 행렬 형태로 표현	- 산점도의 점/마크에 - 여러 의미를 부여하여 - 확장된 차트	- 자료분포 형태를 직사각형 형태로 시각화 - 특정 변수에 대한 구간별 빈도수

=====

3) 비교 시각화

(1) 비교 시각화

- 다변량 변수 데이터를 제한된 2차원에 효과적으로 표현
- 비교 시각화 유형: 플로팅바차트/히트맵/체르노프페이스/스타차트/평행좌표그래프

플로팅 바 차트	히트맵	체르노프 페이스	스타 차트	평행 좌표 그래프
Floating Bar Chart	Heat Map	Chernoff Faces	Star Chart	Parallel Coordinates
				
- 최소~최대 수치까지 막대가 걸쳐있게 표현 - 범주 내 값의 다양성, 중복 및 이상값 파악 - 간트차트라고도 부름	- 여러 변수를 비교 - 각 변수는 열 방향 - 칸 별로 색상 구분 - 최솟값 옅은 색 - 최댓값 진한 색	- 얼굴 하나로 표현 - 데이터를 눈코입 등과 일대일 대응	- 각 변수의 표시 지점을 연결하여 별모양 도형으로 나타냄 - 최솟값은 중심점 - 최댓값은 가장 먼 끝점	- 다변량 데이터를 평면에 가시화

=====

4) 인포그래픽

(1) 인포그래픽(Infographics)

- 중요 정보를 하나의 그래픽으로 표현함
- 복잡하고 어려운 정보/데이터를 쉽고 명확하게 이해할 수 있게 함
- 그래픽과 텍스트가 균형을 이루도록 조합
- 인포그래픽 유형: 지도형/도표형/스토리텔링형/타임라인형/비교분석형/만화형
- 활용방법: 퍼블릭 데이터 활용/템플릿과 아이콘 배치/무료 툴 활용/저작권 설정/인포그래픽스 홍보

2.3 분석 결과 활용

-	Keyword
분석 모형 전개	분석모형전개, 운영시스템적용, 운영및개선방안, 예측오차, 추적신호
분석 결과 활용 시나리오 개발	활용분야분류, 초기아이디어개발관점, 마인드맵방식, 친화도표방식, 피라미드방식, 가치사슬관점, 서비스모형, 채널시스템활용, 비즈니스모형활용, 비즈니스모델캔버스
분석 모형 모니터링	성능모니터링, 모니터링솔루션, 샤이니(R), 분석주기, 성능이벤트, 임계치
분석 모형 리모델링	리모델링

=====

1) 분석 모형 전개

(1) 빅데이터 모형 운영 시스템 적용방안

- 모형개발&운영 -> 적용방안 -> 적용단계
- 빅데이터 모형 개발 및 운영:
분석목적 정의 -> 가설검토 -> 데이터 준비&처리 -> 모델링 및 분석 -> 정확도&성능평가 -> 운영

분석목적 정의	가설검토	데이터 준비&처리	모델링 및 분석	정확도&성능평가	운영
-문제가 무엇인지 방식/목적 명확히	-문제해결에 적합한 가설수립 -통계적으로 유의미한 결론의 활용 검토	-전처리: 변수 정의, 일관성점검, 정제 등 -변수변환/파생변수 -데이터 분할	-검토된 사안을 통계적 질문으로 변환 -학습데이터셋으로 분석모형 도출	-검증데이터셋으로 분석모형을 평가	-분석모형을 운영 시스템과 통합/활용

- 빅데이터 모형의 운영 시스템 적용방안
 - 도출된 모형을 의사결정에 활용하기 위해서 운영 시스템과의 통합이 필요함
 - 운영 시스템과 모형에 사용된 언어가 같은 경우 -> 통합과정 쉬움
 - 통계 패키지로 모형이 개발된 경우 -> 대부분의 경우, 통합과정 어려움 -> 호환을 위해 인터페이스 필요
 - 인터페이스를 통해 모듈 구동 -> 결과값 도출 -> 인터페이스로 직접 전달 or 파일/DBMS로 간접 전달
 - 인터페이스: 서로 다른 2개의 시스템/장치간 정보/신호를 주고 받는 경우의 접점이나 시스템
- 빅데이터 모형의 운영 시스템 적용단계
 - 분석 모형 적용 모듈 결정 -> 통합방식 결정 및 구현

(2) 빅데이터 모형의 운영 및 개선방안 수립

- 예측 오차 계산 -> 예측모형의 점검 여부 결정 -> 개선 방향 결정
- 개선방안 수립
 - 모형을 운영 시스템에 적용 -> 시간이 지남에 따라 성능저하 발생할 수 있음 -> 지속적인 성능추적이 필요함
 - 예측모형에 대한 성능추적: 추적 신호 값을 활용하여, 예측 오차가 계속 증가 or 감소하는지 확인 필요
- 추적 신호(TS; Tracking Signal) = (예측 오차들의 합)/(예측 오차 절댓값들의 평균)
- 모형 개선방안 수립 절차

예측 오차 계산	예측모형의 점검 여부 결정	예측모형 개선 방향 결정
모형을 실행할 때마다 예측오차 계산&기록	예측오차로 추적신호 다시계산 → 관리도를 활용하여 추적신호 추적	<개선방법1> 최근 데이터&같은 모형 예측모형을 업데이트하여 다시 적용
예측오차 = 모형 결과값 - 실제값	관리도 상한 or 하한을 벗어나면 점검필요	<개선방법2> 다른 모형 상호 비교평가를 통해 선정된 모형으로 교체

=====

2) 분석 결과 활용 시나리오 개발

(1) 분석 결과에 따른 활용 분야 분류

- 분석 결과를 활용할 수 있는 분야 파악
 - 직접(1차) / 파생(2차) 활용할 수 있는 업무와 가치사슬 파악
 - 직접 활용: 해당 업무의 가치사슬에서 직접 활용 -> 활용방안이 정의되어 있음
 - 파생 활용: 활용방안 확대 or 유사/관련있는 업무의 가치사슬에서 분야 파악
- 분석 결과를 활용할 수 있는 분야 분류
 - 초기 아이디어 개발/가치 사슬 관점 -> 관련있는 아이디어를 그룹화하여 분류
 - 초기 아이디어 개발 관점의 분류
 - 마인드맵/친화도표/피라미드 방식의 분류
 - 마인드맵: 줄거리를 이해하며 정리
 - 친화도표: 관련있는 아이디어끼리 묶음
 - 피라미드: 계층 구조가 중요한 경우에 적용
 - 가치 사슬 관점의 분류
 - 수평적/수직적으로 통합/확대하여 새로운 가치사슬을 발견
 - 가치사슬: 기업에서 경쟁전략을 세우기 위해, 자신의 경쟁적 지위를 파악하고 이를 향상시킬 수 있는 지점을 찾기 위해 사용하는 모형

(2) 분류 결과를 토대로 적용 가능한 서비스 영역 도출

- 직접 활용 / 파급 활용(융합 활용)

(3) 분류 결과를 토대로 적합한 신규 서비스 모형 도출

- 새로운 서비스 모형을 위해 필요한 작업
 - 모형에 대한 개념 도출 -> 신규 서비스 모형 정의
- 신규 서비스에 대한 사용자와 제공가치 도출
 - 초기 아이디어 개발/서비스 품질/일반적인/시장 전체 관점
 - 개념 도출을 위해서, 사용자와 제공가치를 찾는 것이 가장 중요함

초기 아이디어 개발 관점	서비스 품질 관점	일반적인 관점	시장 전체 관점
모형 도출	제공 가치	제공 관점	고객 분류
제공가치를 통해 사용자 정의 → 기존 모형과 사용자는 동일, 제공가치는 달라질 수 있음	서브퀄 모형 기준 (SERVQUAL)	질적인 가치 성능/ 디자인/ 맞춤형/ 브랜드가치/ 참신성	- 매스마켓형: 세분화X - 틈새시장형: 특화/전문화 - 세그먼트형: 명확히 정의 - 복합 세그먼트형: 그룹 혼재 - 멀티사이드형: 그룹 2개 이상
사용자 정의 후, 제공가치 정의 → 사용자,제공가치 둘다 달라질 수 있음	반응성/ 공감성/ 혁신성/ 유형성/ 신뢰성	양적인 가치 가격/ 편리성/ 효용성/ 접근성/ 무게	

- 신규 서비스 모형 정의: 개념도 / ITO 프로세스 관점
 - 서비스 개념도(청사진) 관점
 - 서비스 명칭/개념설명/사용자/제공가치/주요기능 등에 대해 정의서 작성

- ITO 프로세스 관점
 - 개념도 관점보다 구체적으로 접근 가능(ex. 시스템/투입요소/변환과정/산출 등)

(4) 서비스 모형에 따른 활용 방안 제시

- 채널 시스템 / 비즈니스 모형 활용 방안
- 조직 내부에서 빅데이터 서비스 제공을 위한 채널 시스템 활용 방안 수립
 - 사용자에게 서비스를 제공하는 채널 시스템을 정의함
 - 기존 시스템에 새롭게 기능 추가 or 신규 시스템 개발
- 사업화를 추진하기 위한 비즈니스 모형 활용 방안 수립
 - 수익 창출에 활용할 수 있는 비즈니스 모형
 - "서비스 모형의 사용자"와 "비즈니스 모형의 사용자"가 동일한지 파악
 - 신규 빅데이터 비즈니스 모형 정의
 - 최초 도입 or 기존 모형을 확장
 - 최초 도입인 경우: 핵심 성공 요인, 주요 실패 요인을 파악해야 함
 - 신규 빅데이터 비즈니스 모형 사례: "비즈니스 모델 캔버스"

빅데이터 비즈니스의 핵심 성공 요인	빅데이터 비즈니스의 주요 실패 요인	비즈니스 모델 캔버스
<p><목적 및 참여요소 측면></p> <ul style="list-style-type: none"> - 분석목적/사용자/활용목적 명확하게 정의 - 데이터볼륨보다 <u>가치창출관점</u>에서 검토 - <u>전문가</u> 참여 필수적 <p><프로세스 측면></p> <ul style="list-style-type: none"> - 분석 인프라 요건 검토 - 주기적으로 모니터링 및 정제 - 작은규모로 시작 → 성공사례 공유 및 확장 	<ul style="list-style-type: none"> - 분석목적/ 서비스목적의 불명확 - 사용자/ 활용방안의 불명확 - 분석 대상 데이터 품질의 저하 - 모형에 대한 정의없이 인프라 우선도입 (ex. 보유한 데이터는 비정형인데 정형 분석 솔루션만 있음..ㅠ) 	<ul style="list-style-type: none"> - 빅데이터 분석모형은 목적을 명확하게 정의하는 것에서 성패 좌우됨! - <u>소규모</u>로 시작하여 (Start Small) <ul style="list-style-type: none"> → 성공 경험을 공유하고 → 영역 <u>확대</u>하는 것이 바람직! - 분석결과가 타당해야 서비스개발 가능~

=====

3) 분석 모형 모니터링

(1) 분석 모형 모니터링

- 실시간/배치(일괄) 스케줄러 실행 -> 주기별로 자동 모니터링 -> 이상 시에만 확인
- 분석 모형의 성과가 예상했던 수준으로 나오고 있는지 모니터링
- 모니터링 솔루션: 자체 상태/정상 작동상태 유무/데이터 처리 및 분석 소요시간/모델에 따른 처리성능 관점에서 모니터링 수행
- R Shiny(샤이니)
 - 모델링 결과를 간단히 배포 가능
 - 구성: 사용자 작업 파일(ui.R) / 서버 파일(server.R)
 - 해당 URL에 접속하면, R로 개발한 분석 모델 실행 가능

(2) 분석 모형 성능 모니터링

- 분석 주기/응용 프로그램/응용 플랫폼/응용 솔루션/성능 이벤트/임계치
- 성능 모니터링: 측정 항목 정의 -> 모니터링 실시 -> 측정 항목별 임계치/이벤트 등급별 알람 -> 성능 관리
- 분석 주기별 모니터링 기준
 - 일간/주간/월간/분기/연간
 - 월간/분기: 성능 추이 집계 분석/현황 보고 등
 - 연간: 연간 업무 계획/연간 리포트 등
- 측정 항목별로 영향을 미치는 요소 -> 응답시간/사용률/가용성/정확성

측정항목	응답시간	사용률	가용성	정확성
설명	서비스 요청 시점 ~ 사용자 응답 시점	일정 시간 동안 자원을 <u>정상적으로</u> 사용한 비율	서비스 장애 없이 <u>정상적으로</u> 지속하여 제공할 수 있는 능력	처리 결과 정확성에 영향을 주는 요인
영향 요소	- 정보시스템 처리 성능 - 네트워크구간 처리용량 - 정보시스템 자원 용량	- 네트워크 자원을 일정시간 사용하는 정도	- 하드웨어 장애 - 소프트웨어 버그 - 운영자/전기적 문제 - 장비/서비스 가용성	- 잘못된 환경 설정 - 하드웨어 장애 - 데이터 이상값

- 주요 성능 측정 항목
 - 응용 프로그램 성능 측정 항목: 응답시간/트랜잭션 처리량/메모리 사용/데이터베이스 처리/오류 및 예외/배치 실행 환경
 - 응용 플랫폼 성능 측정 항목: 응답시간/트랜잭션 처리량/대기 큐/대기 시간/프로세스(스레드) 상태 및 개수/세션 상태 및 개수/통신 큐&채널 상태/자원 풀/오류 및 예외/부하 분산
 - 응용 솔루션 성능 측정 항목: 구간별 수행시간/대기 큐/메모리&버퍼/오류 및 예외
- 성능 모니터링 이벤트 유형
 - 성능 이벤트: 설정한 임계치가 초과되는 것
 - 임계치(Threshold)
 - 성능 모니터링을 위해 정의해놓은 측정 항목마다 임계치를 설정함
 - 성능 모니터링 시, 장애 상황 및 성능 상태의 경계선
 - 임계치에 따른 등급 설정 -> 정상 상태를 기준으로 설정
 - 비정상적인 상황을 판단하는 경계
 - 임계치 설정 및 관리
 - 각 구성요소의 특성에 따라 별도로 임계치 설정
 - 임계치 설정: 제공하는 서비스 형태&시스템 특성 고려 -> 사용자 응답시간&처리속도&만족도 등을 반영
 - 임계치 관리: 단일기준X -> 다양한 요소를 반영하여 조정 -> 운영하면서 주변 요소의 영향을 받아 조정
 - 주요 성능 저하 요인: 서버 자원 부족/성능 조정 부족/I/O 조각화 현상/데이터 이동/프로그래밍 오류/데이터베이스 설계 오류/악성코드/버그/하드웨어 다운/외부적 요인 등

(4) 분석 모형 모니터링 고려사항

- 실제 운영시스템에 적용 -> 상용/오픈소스 도구 활용 or 자체개발 -> 데이터크기&처리속도 고려하여 적용

- 상용/오픈소스 도구에서 기능을 제공할 때만, 자동화 적용 -> 모델 적용 및 갱신 자동화
- 기법에 따른 고려사항
 - 시뮬레이션: 모델적용을 위한 프로세스&업무규칙이 문서화되고 공유됨
 - 최적화: 결과를 시스템과 인터페이스 가능하도록 데이터베이스 연동 프로그램을 개발

=====

4) 분석 모형 리모델링

(1) 분석 모형 리모델링

- 리모델링
 - 빅데이터 모형의 지속적인 성과 모니터링을 통해 편차가 일정수준 이상으로 지속적으로 하락하는 경우에 기존 모형에 데이터마이닝/시뮬레이션/최적화를 적용하는 개조작업
- 리모델링 수행주기: 분기/반기/연 단위가 바람직함(일/주 단위는 특수분야만)
- 리모델링 업무 및 주기

기법	데이터 마이닝	시뮬레이션	최적화
리모델링 시 수행하는 업무	동일한 데이터에 대해 재학습 or 변수추가	이벤트 발생 패턴 변화, 시간 지연 변화 등을 처리	목적함수의 계수 변경 or 제약조건의 제약 값 변화&추가
주기	분기	반기 or 주요변경이 이루어지는 시점	연 단위

(2) 분석 모형 리모델링 절차

- 개선용 데이터 수집&처리 -> 분석 모델 개선 -> 분석 결과 평가&분석 모델 등록
- 개선용 데이터 수집 및 처리: 기존 모델 성능 검토/개선 데이터 선정
- 분석 모델 개선: 분석 알고리즘 선정/알고리즘 수행 및 분석결과 기록

개선용 데이터 수집&처리	분석 모델 개선	분석 결과 평가 & 분석 모델 등록
<기존 모델 성능 검토> 현황 분석 → 성능 검토 → 개선필요성 결정	- 모델 개발했을 때와 같은 절차 - 기존 모델보다 성능이 높아지도록 <u>파라미터를 조정하여 다시 개발</u>	- <u>다양한 이해관계자</u> 가 모여 리뷰&선정 - 분석가/ 데이터처리자/ 고객 등
<개선 데이터 선정> 제외할 데이터가 있는지 점검, 정제, 변환	<분석 알고리즘 선정> 개선 목적&데이터 선정 기존 데이터의 변경내역 조사	<평가 및 등록 절차> 평가 기준 선정 → 분석 결과 검토 → 알고리즘별 결과 비교
<개선 데이터 선정 시 고려사항> <u>데이터 활용도/ 변경도/ 오류율/ 데이터 오류율/ 분석가의 판단..</u>	<알고리즘 수행 및 분석결과 기록> 학습/검증/시험용 데이터를 분할할 때, 추가된 신규 데이터가 반영될 수 있도록!	

(3) 분석 모형 리모델링 고려사항

- 정기적인 재평가와 모형 재조정(필요한 경우)이 필요함
- 모형 재조정 주기 설정
 - 초기에 자주 수행(주기 짧게) -> 점진적으로 주기를 길게 설정(업무 특성에 따라 차이 있음)
- 업무 자동화: 관리해야하는 모델이 월 20개 이상 or 다른 업무와 함께 수행해야 하는 경우에 권고됨
- 리모델링 고려사항
 - 데이터 마이닝: 최신 데이터 적용/변수 추가
 - 시뮬레이션: 업무 프로세스 KPI 변경/주요 시스템 원칙 변경/발생 이벤트 건수 증가에 따른 성능평가와 재조정
 - 최적화: 조건 변화/가중치 변화/계수값 조정/제약조건 추가