2과목 데이터분석 기획

분석 대상과 방법

Known UnKnown

Known 최적화(optimization) 통찰(Insight)

UnKnown 솔루션(Solution) 방안(Discovery)

분석기획 방안

과제 중심적 접근 장기적 마스터 플랜

목적 빠르게 해결 지속적 분석 원인 해결

1차목표 speed & Test Accuracy & Deploy

과제유형 Quick & Win Long Tern View

접근방식 Problem Solving Problem Defintion

의사결정을 가로막는 요소

프레이밍 효과: 동일 상황임에도 개인의 판단, 결정이 달라짐

고정관념, 편향된 생각

분석 방법론의 구성요소

절차, 방법, 도구와 기법, 템플릿과 산출물

분석 방법론의 모델

폭포수 모델: 이전단계 완료되어야 다음 단계 진행

나선형 모델: 여러 개발 과정을 거쳐 점진적으로 완성, 위험요소 제거

프로토타입 모델: 일부분을 우선개발 하고 보안

애자일: 일정한 주기를 가지고 프로토타입을 끊임없이 수정하여 고객의 니즈 반영

KDD분석 방법론

데이터선택 -> 전처리 -> 변환 -> 마이닝 -> 결과평가

전처리: 이상값, 잡음 식별 및 데이터 가공

변환: 변수선택 및 차원축소

Crisp-DM 분석방법론

업무이해 -> 데이터이해 -> 데이터 준비 -> 모델링 -> 평가 -> 전개

(업데이트모델평가전)

모델링 단계에서 모델평가 수행하고 평가 과정 단계에서 모델 적용성 평가 수행

평가 -> 전개에서 위대한 실패 발생가능

빅데이터 분석 방법론 순서 및 단어 중요

디자인 싱킹

사용자에 공감으로 시작해서 아이디어 발산/수령 과정을 통한 피드백으로 발전하는 과정

공감하기 -> 문제정의 -> 아이디어 도출 -> 프로토타입 -> 테스트

지도학습, 비지도 학습

1)지도학습: 정답이 있는 데이터를 학습

분류분석, 회귀분석, 의사결정트리 ,KNN, SYM

2)비지도학습: 정답이 없는 데이터를 학습

군집분석, 자원축소, 연관규칙분석

분석 과제에서 고려해야할 5가지 요소

데이터크기, 속도, 데이터 복잡도, 분석 복잡도, 정확도/정밀도

프로젝트 관리 지식 체계 10가지

통합, 범위, 시간, 원가, 품질, 인적자원, 의사소통, 리스크, 조달, 이해관계자

(이범통이 의자에서 시원한 조리품을 먹었다)

우선순위 선정

전략중요도: 전략적 필요성, 시급성

실행용이성: 투자 용이성, 기술 용이성

ISP

중장기 마스터 플랜을 수립하는 절차

분석 거버넌스 체계 구성요소

조직, 프로세스, 시스템, 데이터, 분석관련 교육 및 마인드 육성체계

(시조프로마인드데)

데이터 분석 수준 진단

분석 준비도

1. 분석업무파악: 사실분석, 예측
2. 분석 인력 및 조직: 분석전문가, 관리자
3. 분석기법: 적합한 기법 사용, 라이브러리/평가/개선
4. 분석데이터: 데이터관리, 외부데이터 활용
5. 분석문화: 의사결정, 회의에서 활용, 공유 및 협업
6. IT인프라: 운영시스템 통합, 환경

(IT문데기인파)

분석 성숙도

CMMI모델 기반

비즈니스/ 조직,역량/ IT부문 관점으로 구분

1. 도입: 환경 시스템구축
2. 활용: 업무에 적용
3. 확산: 전사 차원관리, 공유
4. 최적화: 혁신, 성과향상에 기여

(도활확최)

데이터 거버넌스 구성요소

원칙, 조직, 프로세스

(원조프)

조직 및 인력방안 수립

집중구조: 독립적인 전달 조직구성(중복 업무 가능성 존재)

기능구조: 해당 부서에서 직접 분석

분산구조: 분석조직 인력을 현업 부서에 배치

(집기분)

분석주제 유형 및 특징

최적화(Optimization) 대상 분석 방법 이해, 최적화된 형태로 과제 수행

솔루션(Solution) 방법을 알지 못하는경우 솔루션 찾는 방법으로 수행

통찰(Insight) 기본 분석 방식을 활용하여 문제의 도출및 해결에 기여함

발견(Discovery) 방법을 모른다면 발견을 통해 대상 자체를 새롭게 도출함

분석 방법론

1. 기업에 정착하기 위해 데이터 분석을 체계화하는 절차, 정리된 데이터 분석 방법론 필요
2. 구성요소엔 상세한 절차, 방법, 도구와 기법, 템플릿과 산출물이 있다.
3. 기업의 합리적 의사결정에 장애가 되는 요소엔 고정관념, 편향된 생각, 프레이밍 효과 가 있다

KDD 분석 방법론(Knowledge Discovery in Database)

5단계 데이터셋 선택 – 데이터 전처리 – 데이터 변환 – 데이터 마이닝 – 데이터 평가

CRISP-DM 4단계

단계 – 일반화 태스크 – 세분화 태스크 – 프로세스 실행

CRIS-DM 프로세스 6단계

업무이해, 데이터 이해, 데이터 준비, 모델링, 평가, 전개

햐양식 문제검토 순서

문제탐색(Problem Discovery) – 문제정의(Problem Definition) –

해결방안 탐색(Solution Search) – 타당성검토(Feasibility Study)

분석 프로젝트 5가지 특성 관리 영역

Data Complexity 분석하고자하는 데이터 양을 관리 방안 수립 필요

Speed 시스템에 산정되어있는 데이터들을 통합해서 분석 프로젝트를 진행할때 고려

Analytic Complexity 정확도, 복잡도는 트레이드 오프 관계가 필요

Accuracy & Precision 정확도는 분석의 활용적인 측면 정밀도는 분석의 안정적인 측면

데이터 분석 수준 진단 개요

1. 데이터 분석 수준 진단으로 데이터 분석 기법을 구현하기 위해 무엇을 준비하고 보완 해야하는지 알 수 있다
2. 분석의 유형 및 방향성에 영향을 준다
3. 분석 준비도와 분석 성숙도를 함께 평가함으로써 수행 가능

데이터 분석 준비도 6가지

분석 업무파악, 인력 및 조직, 분석 기법, 분석 데이터, 분석 문화, IT인프라