

데이터로 그리는 새로운 세상, 데이터솔루션 +

국내 프로젝트 사례집

www.datasolution.kr

국내 프로젝트 사례집

Contents +

제조 (Manufacturing)

금융 (Banking & Financial Service)

물류 & 유통 (Logistics & Merchandising & Distribution)

엔터테인먼트 (Entertainment)

통신 & 인터넷 (Telecommunication & Internet)

공공 분야 (Public Sector)

HR (Human Resources)

국내 프로젝트 사례집

제조 (Manufacturing)

- | | | | |
|----|-------------|----|------------|
| 01 | 삼성전자 | 07 | TORAY |
| 02 | SK Advanced | 08 | 현대제철 |
| 03 | LS ELECTRIC | 09 | VOLVO |
| 04 | 현대자동차 | 10 | 대한항공 |
| 05 | 삼성SDI | 11 | 삼성전자(북미지사) |
| 06 | 삼성메디슨 | 12 | 삼양 훌딩스 |

01 | 삼성전자

DX부문 Risk 분석 모델 적용 및 안정화

* DX: 가전제품

프로젝트 기간

1차 사업기간 : 12개월

활용 Tool

Brightics ML, Python,
ARIS PPM

* PPM : Process Performance Manager

배경 및 목적

- 업무 자동화 수준 향상, 데이터 기반 의사결정 지원, 예측을 통한 사전 대응 강화를 위해 Next ERP 시스템 기반 ML 플랫폼이 구성되어 있음
- 그러나 초기 개발 이후 운영 관리 인력의 전문성 부족으로 안정적인 모델 운영 및 원활한 추가 개발이 불가능한 제약이 있음
- 이에 따라 전문 인력을 확보하여 ML 조기 안정화 및 효율적인 운영 전환을 통해 업무 효율성 및 품질 확보를 목적으로 함

주요 모델

- Brightics 서버 운영 및 분석 모델 관리

거래선 재무 건전성	<ul style="list-style-type: none"> · 거래선 정보, 채권 발생/반제, 위험 관리 기준 등을 기반으로 매출 채권에 대한 장기 연체 위험을 점수화 하여 위험 거래선을 사전 감지 · 재무 파트에서 선정한 4개(캐나다, 프랑스, 모로코, 칠레) 법인에 대해 D+7/21 일단위 예측 수행
실행 계획 적정성	<ul style="list-style-type: none"> · 전월 판매 실적 및 금월 RTF(실행계획) 데이터를 활용하여 1~3개월 후 판매 실적 예측 · Time Series 모델 및 Regression 모델의 결과를 Ensemble하여 최종 예측 결과 산출 · 매월 모델 수행 후 Hadoop 서버 내 결과 적재 및 BI 서버 결과 전송 자동화
재무 분야 관련 Risk 시나리오	<ul style="list-style-type: none"> · 업무상 다양한 재무 Risk를 방지하기 위해 정의된 시나리오(업체 경비 비정상 처리 등 월단위 9건, 일단위 3건 예측 운영 중)들에 따라 사전에 탐지하고 대응하기 위한 목적 · 대부분의 모델이 여러 정량/정성적 데이터 및 지표들을 활용하여 2/3차원의 Risk Factor로 차원 축소 및 점수화를 통한 임계치 기반 탐지를 수행
Hand-Carry 상품 자동 발주	<ul style="list-style-type: none"> · 핸드폰 액세서리와 같은 일회성 소모품의 수작업 발주 관리에 따라 발생하는 과잉재고 및 결품 등의 Risk 방지하고 자동 발주 프로세스 구축하기 위한 목적 · 주요 178개의 지점과 225개의 품목에 대해 일단위로 적정 재고량을 예측하여 발주시스템에 자동 적용

- PPM 서버 운영 및 Application 개발 관리

Inbound Risk Monitoring	<ul style="list-style-type: none"> · 제품 생산 및 판매에 따른 프로세스 중 물류 유통상 선박 운송 Risk 탐지 · 선적 및 입고 L/T 산출 및 신속한 대응으로 물류비 최소화 · 선박 운송 과정에서 발생하는 다양한 이벤트 로그에 대한 추적 관리 · 대시보드 기반 시각화를 통해 정량적인 지표로 Risk 관리 효율성 증대
----------------------------	--

02 | SK Advanced

수율 예측 시뮬레이터 개발

프로젝트 기간

사업기간 : 10개월

활용 Tool

Python, Dataiku

배경 및 목적

- 최적 운전 계획 수립 시, Data-Driven 의사결정 지표 부재
- 경험 기반 엔지니어 보유 지식을 정량적 수치로 검증 필요

분석 내용

Dynamic 상관 · 회귀 분석

- 현업 엔지니어가 여러 조건(변수/기간/단위/시차/기법)을 유연하게 선택하며 쉽고 빠르게 분석할 수 있는 툴 개발

XAI 기반 Root Cause 분석

- XAI 기술 중 SHAP 패키지를 Customizing하여 원인 분석이 가능한 툴 개발
- Deep-dive 분석을 통해, 문제의 근원 파악 및 새로운 인사이트 도출 가능

수율 예측 시뮬레이터

- 다차원 AI 모델을 유기적으로 연계한 시뮬레이터 구현
 - 운전 기간 별 모델 세분화(Segmentation)
 - 머신러닝 Boruta 기법 기반 핵심인자 선별 최적화
 - 앙상블 기법을 활용한 예측 모델 개발 및 정교화
 - 시뮬레이션 성능 교차 검증

기대 효과

- 운전 시뮬레이션 결과 기반 중·단기 생산 계획 수립
- 예측 수율 및 스프레드를 반영한 잠재 비용 최소화
- 즉각적인 원인 분석을 통한 공정 운전 이해도 향상



03 | LS ELECTRIC

제품 외관검사 딥러닝 모델 개발

프로젝트 기간

1차 사업기간 : 3개월

2차 사업기간 : 5개월

활용 Tool

Python

배경 및 목적

- 제품 양불 판정을 위한 현 외관 검사 시스템의 한계점 보완 및 정확도 개선 필요
- 외부 솔루션 도입을 지양하고 데이터 수집/분석/관리에 대한 신기술 내재화 목적

분석 내용



이미지 수집 자동화

- 공정 검사 라인에서 촬영한 이미지를 클라우드 저장소에 자동 적재하는 프로세스 정립
- 불량이 발생한 측면의 이미지 우선 수집 (학습용)

불량 유형 정의 및
이미지 라벨링

- 제품의 양불 판정을 위한 불량 유형(외관 검사 항목) 정의
- 모델 학습을 위한 이미지 라벨링 수행
 - 라벨링이란? 검출하고자 하는 영역을 좌표 형태로 지정하고 유형을 입력하는 작업

모델 개발 및 검증

- 객체 탐지(Object Detection) 방법을 활용한 불량 검출 모델 개발
 - Faster RCNN, YOLO, SSD 알고리즘을 기반으로 불량 이미지 학습
 - 글자 인쇄 불량의 경우 문자 인식(OCR) 기법으로 검출
 - 학습한 모델로 불량의 위치를 검출하고 유형을 분류하여 모델 성능 검증

모델 적용

- 공정 검사 라인에 모델을 적용하여 제품 기준의 양불 판정

기대 효과

- 육안 검사를 모델 기반 검사로 대체하여 효율적인 작업 인력 재배치 가능
- 자체 개발함으로써 외부 솔루션 도입 및 유지 비용 절감
- 기존 검사 시스템에 비해 운영/관리 포인트가 적으며 유지보수가 쉽고 빠름

(딥러닝 모델:하나의 모델이 다수의 불량을 검출 vs 기존 시스템:불량 유형별 검출 로직이 각각 다름)

03 | LS ELECTRIC

회로 기판 양불 판정 시스템 구축

프로젝트 기간

1차 사업기간 : 3개월

2차 사업기간 : 6개월

활용 Tool

Python, Microsoft Azure

배경 및 목적

- 검사기 판정 기준 또는 촬영 상 문제 등으로 인해, 불량 중 과반이 가성불량*에 해당
- * 가성불량이란? 육안 검사자의 판단으로는 실질적 양품인데 검사기는 불량으로 잘못 판정
- 검사기 판정결과 보완을 위한 시스템 구축 목적



분석 내용

		수치 데이터 기반 판정 모델	이미지 데이터 기반 판정 모델
하이 브리드 모델	데이터	<ul style="list-style-type: none"> 제품 측정값 (길이, 높이 등) 	<ul style="list-style-type: none"> 제품 촬영 이미지
	알고리즘	<ul style="list-style-type: none"> 머신러닝 분류 알고리즘 – Random Forest, Light GBM 등 	<ul style="list-style-type: none"> 딥러닝 분류 알고리즘 – ResNet101, ResNet152 등
	특징	<ul style="list-style-type: none"> 수치 값 분포 기반 양불 판정 기준 수립 양불 판정 기준을 활용한 가성불량 라벨링 	<ul style="list-style-type: none"> 육안으로 판정 가능한 불량유형 정의 모델 학습을 위한 전체 불량 재라벨링
판정 시스템 구축		<ul style="list-style-type: none"> 하이브리드 모델(수치/이미지 모델 연계) 및 미검 방지 필터 탑재 클라우드 컴퓨팅 디바이스를 도입하여 모델 배포 및 관리 효율성 확보 클라우드 저장소에 데이터 및 예측 결과 자동 적재 시스템 관리자용 대시보드 화면 구현 	

기대 효과

- 검사기 불량 자동 판정에 따른 공정 실행률 향상
- 가성불량 감소로 효율적인 작업 인력 재배치 가능
- 육안 검사자의 주관에 의존하여 일관성이 낮았던 불량 판정 과정 개선

04 | 현대자동차

공정 최적화 및 설비 고장 예측 모델 개발

프로젝트 기간

사업기간 : 7개월

활용 Tool

R, Python, HIVE

배경 및 목적

생산 공정상의 품질 향상을 위한 공정 최적화 및 설비 고장 예측 모델 개발을 통해,
품질 생산성을 제고를 통해 데이터 기반 제조 현장 운영 추진

분석 내용**가공기 툴 부하 모니터링
데이터와 공정 품질 상관성**

- 툴이 받는 부하 데이터를 활용하여 툴의 마모 또는 파손 등 툴의 상태를 예측
- Python으로 개발하여 시스템 적용하여 후 모니터링

**블록 가공 공정품질과
설비 상태 상관성**

- 공정 품질 데이터를 활용하여 제품의 불량 발생 여부를 예측
- 개발 모델의 지속적인 활용을 목적으로 표준화 모델을 개발하여 적용

**PLC 모니터링 DB를 활용한
설비 예방 보전 모델 개발**

- PLC란 제조 사업에서 자동제어 및 감시에 사용되는 제어 장치
- 정해진 시간 안에서 반응하지 않는 특정 동작을 찾기 위해 Gantt Chart를 개발하여 적용

기대효과

- 품질 개선 / 생산성 향상
- 데이터 기반 제조현장 운영체계 지원
- 표준화 모델을 통한 지속적인 운영



04 | 현대자동차

생산공정 데이터 간 상관성 분석

프로젝트 기간

사업기간 : 4개월

활용 Tool

R, Python

배경 및 목적

- 주조공정에서 생산되는 실린더 블록의 불량률 개선
- 공장 비가동 요인 파악 및 최적화를 통한 비가동 기간 단축
- 내구시험기 비정상 정지에 대한 조치 시간 단축

분석 내용

주조공정의 품질 예측 모델 개발

- Outlier Detection 모델을 통한 예측 모델 개발
- 주조 공정의 불량에 영향을 미치는 인자 탐색
- 불량에 영향을 미치는 인자에 대한 관리선 최적화

비정형 데이터를 통한 공장 비가동 분석

- 비정형 데이터의 정형화(향후 다양한 방면에 활용 가능)
- 개체사전 및 동의어 사전을 생성하고 이를 통한 향후 비가동 관련 정보 정형화
- 다양한 관점의 분석을 통한 비가동 요인 탐색

내구시험 비정상 엔진정지 예측 모델 개발

- 시계열 데이터의 분류 모델 개발 방법론 정립

기대효과

공정 최적화를 통하여 생산 품질을 향상시킬 수 있으며, 프로세스 운영 효율화로 운영 비용 절감이 기대됨



04 | 현대자동차

차량 고장 데이터 자동 분류 알고리즘 개발

프로젝트 기간

사업기간 : 6개월

활용 Tool

R, Python

배경 및 목적

차량 이상 데이터 분석 결과 및 노하우의 지식화 및 활용 부재.
따라서 차량 이상 Waveform 데이터 DB화를 통한 지식화/분석 효율화 진행

고장 자동 분류 모델 개발

- 차량 센서 데이터 전처리를 통한 분석 마트 생성
- 시계열 데이터 유사 패턴 탐지 및 선행 이벤트에 의한 후행 반응센서의 변화를 통해 유의 인자 도출
- 자동 분류 알고리즘 적용 및 결과 평가

미분류 차량 데이터 군집화

- 자동 분류가 되지 않은 차량에 대해서 센서 데이터를 통해 군집화를 진행하고, 차량 정비 전문가가 식별할 수 있는 군집 유형을 정의하여 실제 원인을 파악하는데 활용하기 위한 군집화 수행
- 정확도 높은 알고리즘 개발과 고장 유형별 유사도 산출에 따른 수행 속도 최적화



04 | 현대자동차 남양연구소

외판강성 성능 예측을 위한 머신러닝 적용 해석기법 개발

프로젝트 기간

사업기간 : 4개월

활용 Tool

Python

배경 및 목적

새로 개발되는 자동차의 외판강성을 측정하기 위해 과다한 공수가 필요함
M/L 알고리즘을 통해 취약부위를 사전에 예측하여 외판강성 해석공수를 절감하는데 목적이 있음

취약부위 예측 모델 개발

- 9개 부품의 외판강성 시뮬레이션 데이터를 Parsing하여 분석 가능한 데이터로 변환
- 분석 Target을 정의하고 취약부위 예측에 필요한 주요 인자를 정의하여 분석 마트 생성
 - 3차원 좌표 정보를 통해 외판의 곡률을 계산하여 휘어짐 정도를 수치화
 - 용접 부위, 실러 위치, 고정 부위 등을 고려한 주요 인자 생성
- 다양한 M/L 알고리즘을 적용하여 최적의 알고리즘을 선별

취약부위 예측 모델 고도화 자동화

주기적으로 적재되는 시뮬레이션 결과 데이터를 학습 데이터로 변환하여
자동으로 모델을 고도화 하는 프로세스를 개발



05 | 삼성SDI

Cancap3D 용접검사를 위한 딥러닝 모델 개발

프로젝트 기간

사업기간 : 2개월

활용 Tool

SaigeVision, 삼성SDI AI 플랫폼

배경 및 목적

- 자동차 배터리 제조 과정 중 Cancap 용접 불량 검사를 위한 AI 모델 필요성 증대
- 울산 공장 모델을 기반으로 형가리 공장 설비 라인에 대한 모델 최적화 및 적용을 목적으로 함

분석 내용

- Cancap3D 영상을 통한 2D 이미지 수집
- 불량 유형 분류 및 이미지 라벨링
 - Cancap 가장자리 부근에서 발생하는 용접 불량(파이거나 튀어나옴, 미용접 등)을 유형별로 분류
 - 불량 이미지에 대해 영역별 라벨링 수행
- 불량 영역 검출을 위한 Segmentation 모델 개발 및 성능 검증
 - 설비 라인별 과정 및 미검 관련 지표 기반으로 학습 파라미터 최적화

기대효과

- 육안 검사를 모델 기반 검사로 대체하여 인력 및 관리 비용 절감
- 모델을 자체 개발하여 유지보수가 쉽고 빠르며, 설비 적용 및 운영이 용이함



06 | 삼성메디슨

수요예측분석
시스템 구축

프로젝트 기간

사업기간 : 3개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

삼성메디슨이 판매하는 의료 서비스 자재의 효율적 재고관리를 목적으로 자재 수요 예측 모델을 개발함

수요예측분석 시스템 구축

수요예측 Logic 자동화

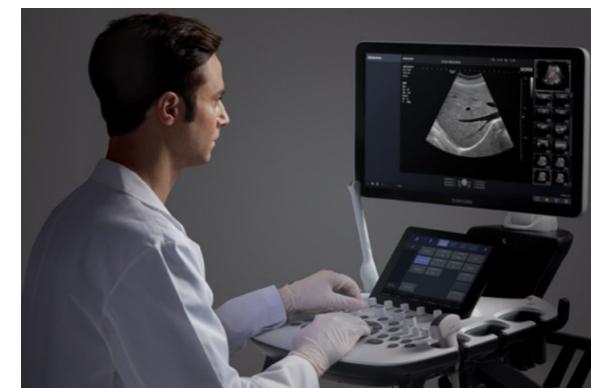
- 자재 타입 별 현업에서 활용 중인 수요예측 Logic 자동화
- 수요예측 값 보정을 위한 참조 값 산출

수요예측 모델 개발

- 자재 수요 빈도와 수요 주기 기준의 Model Segmentation
- 저 빈도 자재에 대한 Zero Logic 개발
- 다 빈도 자재에 대한 Advanced Logic, 회귀모델, 지수평활 모델 개발
- Advanced Logic, 회귀모델, 지수평활 모델 중 최적 모델 추천 시스템 개발

기대효과

적정 수준의 재고를 유지함으로써 과다재고와
재고부족에 대한 운영 비용 손실을 줄일 수 있음



07 | TORAY

특정 공정이상 예측분석

프로젝트 기간

사업기간 : 4개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 필름 제조공정에서 발생하는 공정 이상으로 인해 제조원가가 상승되고 제조 경쟁력이 약화되고 있음
- 공정에서 측정되는 대량의 데이터를 축적하고 있으나 공정 이상과 연계한 분석은 미비함

정상/이상정지 분류 모델 개발

- 의사결정나무, SVM, 로지스틱 회귀분석 등 5가지 분류기법을 적용한 분류 모델링 후 모델 성능을 비교하여 품종 별 최종 모델 확정
- 공정 이상 발생 규칙을 생성하고 주요 요인의 영향도를 도출한 후 공정 정상/이상정지를 예측함

기대효과

- 제조공정 이상의 원인을 도출하고 공정 이상 발생 가능성을 사전에 탐지 가능함
- 데이터 분석 인프라를 구축하고 원가절감 및 제조 경쟁력을 제고함



08 | 현대제철

스크랩 예측 모델 개발

프로젝트 기간

사업기간 : 6개월

활용 Tool

IBM SPSS Statistics &
IBM SPSS Modeler**배경 및 목적**

기존 생산 공정은 관리자 및 엔지니어의 경험으로 관리/운영됨에 따라
데이터에 기초한 통계적 기법 및 기계학습을 활용한 객관적이고 과학적인 모델 개발/검증 필요

스크랩 예측 모델 개발

스크랩 입고 품질 지수 모델	입고 실적 변수를 탐색하고 세부등급에 따른 업체별 입고 품질 점수화 모델 개발
회수율/에너지 예측 모델	선형회귀분석, 의사결정나무분석 등을 적용하여 스크랩 등급 및 조업 지표를 활용한 회수율 예측 모델과 에너지에 영향을 끼치는 인자를 활용한 에너지 예측 모델 개발
전기로 용강 성분 예측 모델	스크랩 등급 및 조업 지표를 활용한 회귀모델 기반의 용강 성분 예측 모델 개발

기대효과

최적의 품질 지수 및 성분 예측으로 생산 공정의
효율성을 증진시키고 비용을 절감함



08 | 현대제철

GQMS

품질분석

기능 강화

프로젝트 기간

사업기간 : 1개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

품질 전문가 육성 및 품질 분석 기능 강화

품질 분석 전문가 육성

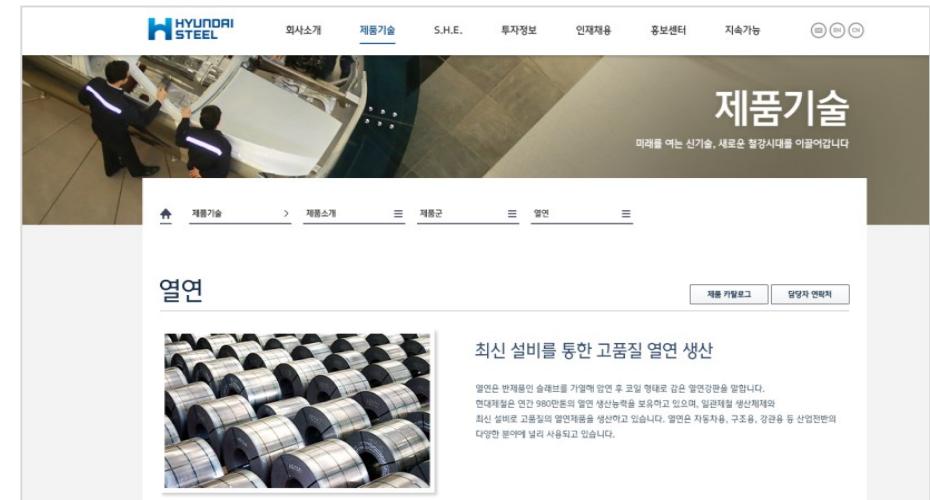
실제 철강 제조 공정 샘플 데이터를 활용한 실습 위주의 교육 및 분석 프로젝트 공동 진행으로
분석의 노하우를 공유하여 전문성과 실용성을 갖춘 품질 분석 전문가를 육성함

머신러닝 기법을 적용한 품질 분석 모델 개발

- 각 기준의 관리도 외에 이상치 탐지와 최근점 이웃법을 추가로 적용하여 품질 이상 발생 패턴을 식별함
- SVM, 의사결정나무, 랜덤 포레스트 등의 머신러닝 기법을 적용하여 제품별 불량률을 예측함

기대효과

IBM SPSS Modeler 분석 전문가를 육성함과 동시에 정형적인 품질 이상 발생 패턴을 식별하고
불량률을 사전에 예측함으로써 품질 분석 기능 강화를 기대함



08 | 현대제철

품질 개선 시스템 구축

프로젝트 기간

사업기간 : 4개월

활용 Tool

IBM SPSS Modeler

배경 및 목적(분석 전용 S/W 도입)

- 기존 Matrix Query와 Excel로써 진행하던 품질 분석 업무를 보다 체계적인 분석을 위해 IBM SPSS Modeler를 전사 프로그램으로 선정
- 결함원인 분석, 품질분석 전문가 육성 및 인프라 구축

품질분석 업무 역량 강화

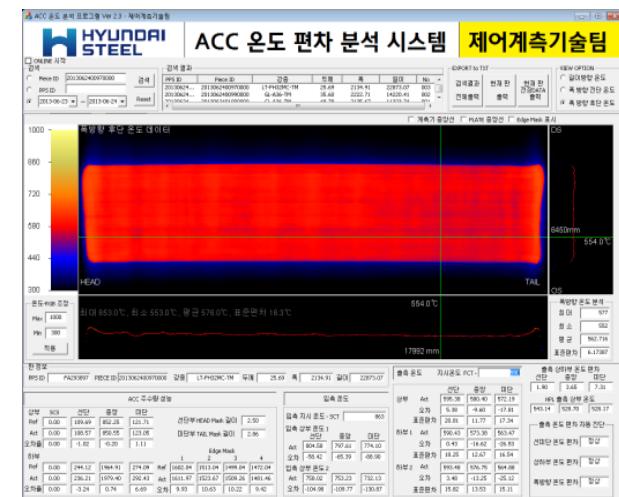
체계적인 분석 프로세스를 확립을 위한 다양한 형태(교육, 세미나, 프로젝트, 기술지원)의 역량 강화 활동 시행

각 업무 별 불량원인 Rule 생성 및 활용

각 부서별 추진과제 모델링 작업 시행 및 일부 공정에서 개선 부분 적용
(에지선상띠흡 부적합 개선 모델, 자외판 에지스캡 부적합 개선 모델, 후판UT 부적합 개선 모델, 고단합금강 부적합 개선 모델, 후판 스키 예측분석 등…)

IBM SPSS Modeler 시스템화

데이터 마이닝 기법을 통해 불량률이 개선된 주제에 있어, 불량 원인 Rule을 탑재한 시스템 구축 지원





09 | VOLVO

건설장비 판매량 예측

프로젝트 기간

사업기간 : 3.5개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

경영계획 수립 시 기존 딜러의 수요예측정보 외에, 통계적 모델 기반으로 예측한 건설장비 판매량을 4C 시스템에 제공함으로써 더 정확한 수요예측정보를 활용하고자 함

건설장비 수요예측 모델 개발건설장비 판매량에
영향을 미치는 요인 탐색

- 데이터 구조 변환 및 전처리
- 분기/Dealer/Machine Line별 판매량 데이터 및 기타 입력변수 EDA

통계적 기법을 통한
건설장비 판매량 예측
모델 개발

- Hub China와 Hub EMEA West 지역의 Dealer-Machine Line별 18개월 판매량 예측
- GLM 기법 및 최근성 가중치를 적용한 모델 개발
- 카이제곱, MAPE 등 평가 지표 선정

건설장비 판매량 예측을
위한 부분 자동화 체계 구축

- 예측 모델 배치 및 Inbound/Outbound 데이터 구조의 자동화 연계

기대효과

보다 정확한 수요예측 정보를 기반으로 건설장비를 생산함으로써, 재고량을 줄이고 건설장비 공급에 필요한 리드 타임을 줄일 수 있음

10 | 대한항공

항공기 이상 징후 탐지모델 개발

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

수작업으로 결함 발생 원인을 파악하는 업무를 통계적 방법론을 이용하여 효율적인 프로세스를 정립하는 것이 필요

이상징후 탐지 모델 개발

- 항공기의 직전 비행에서 발생한 센서 데이터를 이용하여 정상패턴과 비정상 패턴을 식별하는 모델을 개발
- 정상 CASE와 비정상 CASE의 차이를 구분할 수 있는 특성 변수 생성이 중요
- Anomaly Detection & KNN 기법 적용

기대효과

- 문제 발생 징후를 사전에 탐지함으로써 이륙 전 예지 정비를 통해 회항에 대한 손실 감소
- 고장 원인에 대한 특성을 파악함으로써 문제의 정확한 원인을 밝히고 이에 대한 개선이 가능



11 | 삼성전자(북미지사)

VoC 사전 예측 모델개발

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler
Text mining module**배경 및 목적**

비정형(텍스트) 정보에 대한 추출 및 리스크 감지와 서비스 프로세스 내 신규 리스크 감지 요구에 대한 시스템 내로직 확장 및 적극적 고객 대응 체계 수립

CX(고객) Risk Sensing Model

삼성전자 북미지사의 백색가전 구입 고객이 제품 서비스 요청 건으로 서비스 센터에 contact 하는 과정에서 call 인입 단계의 Risk Sensing을 위해 고객 기본정보 및 과거 서비스 프로세스 point 정보를 기반으로 VOC 발생 확률을 예측하여 고객 불만사항을 사전에 탐지, 고객 만족도 제고 및 VOC 발생 방지를 통한 VOC 대응 COST 절감

INN Progress Risk Sensing Model

고객의 보유제품 이상으로 인한 서비스 프로세스 단계에서 발생할 수 있는 VOC를 사전에 예측하여, 고객 불만사항을 사전에 탐지, 적절한 대응 방안 마련을 통해 고객 만족도 제고 및 VOC발생 방지를 통한 VOC대응 COST 절감

TA(Text Analysis for cx comment)

고객 서비스 단계에서 발생하는 고객 불만사항 관련 비정형 데이터(고객 상담내용)를 분석하여, 기존 고객의 제품별 Risk Word를 정의, 이를 기반으로 신규 고객 대응 단계에서 Risk Word 검출 시 이에 대한 적절한 대응을 통해 VOC 사전 예측을 통한 VOC 대응 COST 절감 및 고객 만족도 제고



12 | 삼양 홀딩스

공정 최적화 및 수율 향상을 위한 데이터 분석 모델 개발

프로젝트 기간

사업기간 : 5개월

활용 Tool

Python

배경 및 목적

- 현 제조공정은 작업자 노하우 기반으로 작업조건을 설정하며 표준화된 기준이 없는 경향이 있음
- 공정별 특성을 고려한 데이터 기반의 의사결정 시스템 구축 필요

분석 주제 및 내용

분석 주제	분석 내용
모노파스트 방사공정 작업조건 최적화 (모노파스트 : 수술용 봉합사)	<ul style="list-style-type: none"> · 제품화 소요 시간 단축을 위해 칩 물성에 따른 표준 공정조건 기준 필요 · 공정 조건의 최적 공정 값 탐색을 위한 시뮬레이터 개발 · ANN 모델을 활용한 매듭강력/직선강력 예측 및 최적 공정 조건 제시
밀가루 제분공정의 원맥 목표수분 예측	<ul style="list-style-type: none"> · 수분 규격 이탈로 인한 불필요한 비용 발생 · 공정 내/외부 온습도와 목표수분 간의 인과관계 도출 · Extra tree 모델을 활용한 가수공정 1차/2차 목표수분 예측 모델 개발
MVR 담액 농축공정의 운전조건 최적화 (MVR : 스팀 재활용을 위한 압축 장치)	<ul style="list-style-type: none"> · 물가 상승으로 인한 운영 비용 증가 및 설비 가동 효율성 상승의 필요성 증대 · 농축 후 Bx에 영향을 미치는 주요 변수 선정 및 최적 Rpm 산출 · Decision Tree & XGBoost 를 통한 Rpm오차 주요변수 선정 및 오차값 추정
물엿 농축공정의 공정 작업과정 최적화	<ul style="list-style-type: none"> · 물엿의 최종 Bx를 최적화하여 제품의 Loss를 감소를 목표로 진행 · 인과모델(Linear regression)을 활용한 제품 Bx에 영향을 주는 변수 파악 · 데이터 구간별 모델 생성 및 최적 Bx를 위한 작업조건 규명
전분당 당화공정의 IMO 당화조건 최적화 (IMO : 이소말토 올리고당)	<ul style="list-style-type: none"> · 프로바이오틱스 사업 확대 및 IMO 생산량 증가로 효소 사용량 변동성 증가 · 교차분석, 상관관계 분석 등 IMO 당화공정에서 영향도 높은 변수 선택 및 향후 고도화 방안 제시

국내 프로젝트 사례집

금융 (Banking & Financial Service)

- | | | |
|------------|-----------|------------|
| 01 KB국민카드 | 06 BC카드 | 11 교보생명 |
| 02 DB손해보험 | 07 미래에셋생명 | 12 알리안츠생명 |
| 03 대신증권 | 08 동양생명 | 13 NH농협중앙회 |
| 04 농협생명 | 09 ING생명 | |
| 05 MG새마을금고 | 10 신한생명 | |

01 | KB국민카드

빅데이터 허브 기반 데이터 사이언스 플랫폼 구축(1/4)

프로젝트 기간

사업기간 : 5개월

활용 Tool

HIVE & Impala SQL
R, Python, PySpark 등**배경 및 목적**

- KB국민카드의 기존에 운영중인 VOC 시스템의 데이터 수집 및 처리의 환경적인 제약을 극복하고, 효율적으로 전 KB계열사 데이터 분석 환경 제공을 위한 통합 데이터 사이언스 플랫폼 구축의 필요성
- ※ VOC : Voice of Customer = 소비자 요구를 반영하여 고객관리 (이탈, 연체, 민원 등)

분석내용

개인별 맞춤형
마케팅 추천
스코어링 모델

세그먼트 재 구성 및 신규 세그먼트 발굴

- 기존에 운영하고 있던 100여개의 세그먼트를 고객 단일 관점으로 재 구성함,
- 고객의 니즈 및 현 시점에 맞는 트렌드를 반영하고, GA데이터를 통한 상권분류 등 다양한 방법을 통해 기존 세그먼트 고도화 및 신규 세그먼트 발굴하여 약 350개의 세그먼트 생성

세그먼트 스코어링 모델

- 타겟 마케팅에 활용하기 위해 각 세그먼트에 속하는 고객들에게 해당 세그먼트에 속하는지 여부와 점수와 등급을 부여하여 마케터가 대상자를 추출할 때 효과적인 운용 가능

기대효과

기존 Rule 방식의 세그먼트로 인해 타겟 마케팅을 위한 우선순위 대상자 선별을 통해
고객 만족도와 수익 창출에 기여

01 | KB국민카드

빅데이터 허브 기반 데이터 사이언스 플랫폼 구축(2/4)

프로젝트 기간

사업기간 : 5개월

활용 Tool

HIVE & Impala SQL
R, Python, PySpark 등**배경 및 목적**

- KB국민카드의 기존에 운영중인 VOC 시스템의 데이터 수집 및 처리의 환경적인 제약을 극복하고, 효율적으로 전 KB계열사 데이터 분석 환경 제공을 위한 통합 데이터 사이언스 플랫폼 구축의 필요성
- ※ VOC : Voice of Customer = 소비자 요구를 반영하여 고객관리 (이탈, 연체, 민원 등)

분석내용
**최적 카드상품
추천 모델**
카드상품 혜택 기반 모델 :

- 회원의 이용 실적과 카드상품의 혜택 정보를 고려하여, 회원x카드 별 추정 혜택 값을 산출함. 해당 혜택 값이 가장 높을 것으로 추정되는 카드 상품을 추천

라이프 스타일 소비패턴 기반 모델 :

- 개인별 맞춤형 마케팅 추천 스코어링 모델에서 생성한 고객의 라이프 스타일 영역의 소비패턴을 기반으로, 소비 성향이 유사한 집단의 인기 상품을 고객에게 추천

회원 속성 정보 기반 모델 :

- 신용카드 이용정보가 없는 무실적 회원과 실적 정보가 있는 체크카드 회원을 대상으로 적합한 신용카드를 추천하기 위해, 회원의 공통적인 속성 정보를 활용하여 추천

대체 카드 추천 모델 :

- 카드 상품 혜택 정보를 기반으로, 카드상품 간의 유사도 분석을 통해 관계를 파악하고, 이를 활용하여 운영 중단 카드를 보유한 회원에게 대체할 수 있는 카드를 추천

기대효과

막연하게 카드를 사용하는 고객에게 혜택기반 및 라이프 스타일 기반의 카드를 추천해줌으로써, 충성도 향상 및 이벤트 효과 기대

01 | KB국민카드

빅데이터 허브 기반 데이터 사이언스 플랫폼 구축(3/4)

프로젝트 기간

사업기간 : 5개월

활용 Tool

HIVE & Impala SQL
R, Python, PySpark 등**배경 및 목적**

- KB국민카드의 기존에 운영중인 VOC 시스템의 데이터 수집 및 처리의 환경적인 제약을 극복하고, 효율적으로 전 KB계열사 데이터 분석 환경 제공을 위한 통합 데이터 사이언스 플랫폼 구축의 필요성
- ※ VOC : Voice of Customer = 소비자 요구를 반영하여 고객관리 (이탈, 연체, 민원 등)

분석내용

가맹점 단위
마케팅 타겟
정교화 모델

- 가맹점(요식업) 관점에서 방문 가능 고객 또는 단골 고객을 추정하는 가맹점 관점 고객 스코어 모델 개발.
- 단골 가능 스코어 모델 :
 - 가맹점 및 이용 고객의 특성을 기반으로 가맹점 별 고객의 스코어 산출
- 방문 가능 스코어 모델 :
 - 가맹점이 속한 업종/상권에서 잠재고객의 결제 특성을 활용하여 산출
- 특화 스코어 모델 :
 - 주중/주말 시간대 별 특성과 상권, 업종 요인을 고려하여 산출하고, 타겟 마케팅 시나리오에서 활용

고객 VOC
통합 모델

- 이탈 예측 모델 :
 - 모델 재설계를 통해 이탈 고객에 대한 예측력 증가 및 유지보수 효율성 증대
 - 연체 예측 모델 :
 - 고객별 이용실적과 비즈니스 요구 반영한 모델 설계로 예측력 확보
 - 민원 예측 모델 :
 - 민원 내용을 분석하여 입력변수로 활용하여 최적화를 통한 운영 효율 증대
- ※ VOC 모델은 비정형 데이터를 활용하여 예측성능을 더 높임.

기대효과

- 가맹점 관점의 모델을 활용하여, 가맹점과 파트너를 유지할 수 있고 가맹점별 매출실적을 상승시킬 수 있음
- VOC 모델을 현재 트렌드에 맞게 최신화하고 신규 Insight 도출을 통해 고객 관리의 효율성 증대

01 | KB국민카드

빅데이터 허브 기반 데이터 사이언스 플랫폼 구축(4/4)

프로젝트 기간

사업기간 : 5개월

활용 Tool

HIVE & Impala SQL
R, Python, PySpark 등

배경 및 목적

- KB국민카드의 기존에 운영중인 VOC 시스템의 데이터 수집 및 처리의 환경적인 제약을 극복하고, 효율적으로 전 KB계열사 데이터 분석 환경 제공을 위한 통합 데이터 사이언스 플랫폼 구축의 필요성
- ※ VOC : Voice of Customer = 소비자 요구를 반영하여 고객관리 (이탈, 연체, 민원 등)

분석내용

비정형 텍스트 분석

트리거 모델링 :

- 성능과 속도를 개선을 통해 비즈니스 목적에 따라 대상군과 대조군을 분석하고 주요 키워드를 정의하여 매월 최신화 된 트렌드가 반영된 동적 모델을 개발

개인화 추천 모델

기존 개인화 추천 모델의 데이터 수집부터 처리하는 프로세스를 최적화 진행

※ 개인화 추천 모델의 가장 큰 이슈였던 처리 속도를 획기적으로 단축함.

기대효과

모델의 처리 속도를 개선하여 운영 관점에서의 분석모델 관리 효율성 증대

02 | DB손해보험

일반보험 UW 예측 모델링 및 시스템 고도화

프로젝트 기간

사업기간 : 10개월

활용 Tool

IBM SPSS Modeler
다빈치 Labs

배경 및 목적

- UW의 직관적인 판단으로 Risk평가의 정확성이 높지 않고, 일관성 없는 심사결과에 대한 현장의 불신이 증가
 - Risk 평가범위를 확대하고, 과학적으로 정량화된 Risk 산출을 통해 손해율 개선과 업무 효율화가 목표
- ※ UW : Underwriting = 보험 계약 건에 대해서 개인의 발생가능한 위험이나 손실의 정도를 결정하는 과정(부서)

분석내용

손해액
예측 모델

- 일반보험 상품별 포트폴리오(모델링 대상 세그먼트)를 30개로 구성함
 - 일반보험의 손해액 및 손익 예측분석 프로세스 수립
 - 빈도와 심도를 타겟으로 예측모델 개발
 - 심사과정에서 발생하는 Factor의 위험도 적용
 - 머신러닝 알고리즘과 GLM 기법을 활용한 최초 모델 구축
 - 손해율 추이를 탐색하여 조기에 경보를 발생해주는 모델 개발
- ※ GLM : Generalized Linear Model(일반화 선형 모델)

기대효과

보험을 심사하는 심사자별 심사건수가 30% 감소하고, 심사되는 건 별 심사시간 20%를 단축하여 효율성 증대

02 | DB손해보험

UW 리스크 관리 시스템 구축

프로젝트 기간

사업기간 : 15개월

활용 Tool

IBM SPSS Modeler
Client/Server

배경 및 목적

- UW관점에서 위험요인을 UW전용 데이터베이스에 최소 5년 이상 축적하고 적시에 분석이 가능한 리스크 관리 시스템을 구축

※ UW : Underwriting = 보험 계약 건에 대해서 개인의 발생가능한 위험이나 손실의 정도를 결정하는 과정(부서)

분석내용

Risk Factor Pool 구축

- UW전용 Data Warehouse에 5년 이상의 위험요인 데이터를 축적 및 관리하여 즉시 필요한 시점에 분석을 진행할 수 있는 기반 마련

※ Data Warehouse : 기간 시스템의 데이터베이스에 축적된 데이터를 공통의 형식으로 변화해서 관리하는 데이터베이스를 의미.

Scoring 시스템 구축

- 일반보험종목 스코어 모델을 손해보험사 최초로 개발 및 구현
- 장기보험종목 스코어 모델을 고도화하고 역선택 유형을 감지하는 모델 손해보험사 최초 적용
- 운영 스코어 모델 관련하여 성능 평가 및 리모델링 기능을 시스템화를 통해 팩터(영향요인) 변경 및 상대도 변경 시 즉시 모델 Update 기능 구현

조기 경보 분석

- 상품/PA/계약 정보를 바탕으로 최근시점 기준 손해율 변동을 감지하여 현 시점에 적합한 위험 그룹을 식별하는 모델 구현
- 심사화면에서 바로 정보를 확인할 수 있도록 함

※ PA : Prime Agent = 보험 컨설턴트

기대효과

- 위험평가의 정교화 및 심사차별화를 통해 UW 리스크 관리 기반을 구축하고, 일반/자동차 Rule도입으로 현업관리 용이
- 명확한 업무 Role에 의한 프로세스별 담당자의 업무분장으로 핵심 심사 업무에 선택과 집중 가능
- UW각종 통계자료의 계층별 제공으로 인수 정책 수립 및 현업의 계약인수 결정의 적시성 제공

03 | 대신증권

신온라인 고객관리 시스템 구축

프로젝트 기간

사업기간 : 7개월

활용 Tool

IBM SPSS Modeler, R

배경 및 목적

- 모바일 등 플랫폼 변화에 따른 24시간 고객관리 필요
- 신규고객 확보를 통한 양적 증가와 수익성 기반 고객의 질적 증가 필요

분석내용

고객세분화

- 수익성 및 충성도 기준을 정의하여 이를 기반으로 25개의 세분고객군을 도출.
- 통합 및 분류과정을 통해 8개 고객군으로 세분화함.

예측모델링

- 신규/활성/휴면/이탈 고객을 정의하고 마케팅 전략에 따라 이탈방지 모델과 신규 고객 활성화 모델 개발

상품추천

- 고객의 특성 및 행동로그, 과거투자경험, 유사고객군 포트폴리오 등을 분석하여 상품을 추천하고 투자권유 지침에 따라 고객에게 상품 추천 가능 여부 반영

가망고객현황

- 고객의 웹 접속기록, 모의투자 신청여부, 이벤트참여 현황을 파악하여 웹 및 모의투자 채널에 대해 신규고객이 될 수 있는 가망고객으로 정의

기대효과

새로운 금융 트렌드를 선도하고 고객의 니즈를 충족시킬 수 있는 新온라인 고객관리 체계 및 업무 자동화 인프라 구축

04 | 농협생명

유지율 예측

프로젝트 기간

사업기간 : 4개월

활용 Tool

SAS

배경 및 목적

- 보험에 가입한 비율이 한계치에 도달하고 신규 고객수요를 발굴하기 위한 나이도가 증가하는 등 변화하는 환경에 대응하기 위한 마케팅 추진방향 모색 필요
- 성숙기 보험시장에서 지속 성장하기 위해 보유하고 있는 고객 중심의 마케팅 역량 강화 필요

분석내용

신 계약
유지률
예측 모델

- 고객, 계약, 설계사 속성별로 유지율에 영향을 주는 요인을 탐색하여 유의항목을 선정하고, 선정된 유의항목에 대해 구간화를 통해 FC/ AM 채널의 각 4/7/13회차 예측모델을 개발
 ※ FC : Financial Consultant(설계사)
 ※ AM : 대리점

추가판매
예측 모델

- 보험 상품별(보장성/종신/저축/연금)로 로지스틱 회귀분석을 통해 고객의 추가 보험가입 모델을 개발
- 인구통계학적 정보, 거래단계, 거래실적, 생애주기 기준의 고객세분군을 정의하여 마케팅 및 캠페인 진행 시 타겟 고객추출에 활용

기대효과

- 상품x채널 기반의 최적화된 개발을 통해 고객 니즈 충족 및 고객관계관리 강화를 통한 추가계약 확대 기대
- 효율적인 데이터베이스 활용을 위한 프로세스 및 시스템 개선

05 | MG새마을금고

신용평가 시스템 구축

프로젝트 기간

사업기간 : 6개월

활용 Tool

SAS

배경 및 목적

- 현재 운영하고 있는 분석모델의 노후화로 인해 변별력 및 효율성 저하
- 금융감독 당국의 규제 및 권고사항을 고려한 리스크 관리 모델 구축 및 신상품(중금리 상품) 운영전략 필요

분석내용

신용평가 모델 개발

- 새마을 금고 특성을 고려한 세부 분류 기준을 정의한 후 각 요소별 결합으로 최적화된 모델 세분화 기준 결정
- 내부, 대출, 카드, 연체 정보를 기반으로 로지스틱 회귀분석을 적용하여 평점표 방식의 신용평가 모델 개발

기대효과

기존 운영 시스템보다 더 정교화된 신용평가 시스템을 바탕으로 정교한 개인 및 소호(SoHo) 고객 관리

※ 소호 : 1명에서 10명 정도가 일하는 작은 회사를 의미

06 | BC카드

커머스 DB 기반 마케팅 모델 개발

프로젝트 기간

사업기간 : 4개월

활용 Tool

SAS

배경 및 목적

- 내부 데이터베이스를 활용하여 마케팅에 활용할 모델 개발 필요(내부 데이터베이스 활용도 상승 효과)
- 외부에서 수집할 수 있는 데이터(행정동 단위의 금융통계, NICE 제공)를 함께 활용하여 마케팅 대상자를 확대하고 목적에 맞는 모델을 개발하고자 함

분석내용

판매 성공 예측 모델	POM고객을 대상으로, MD/TM 사업, DCDS 사업에 판매 성공률이 높은 가망 고객들을 예측하는 것으로 로지스틱 회귀분석을 활용한 스코어 카드 방법론 적용 ※ POM : Policy Owner Marketing, 기존 보험상품을 보유한 고객 ※ MD/TM : Merchandiser / TeleMarketing ※ DCDS : 채무면제, 유예 상품
Call 성공 예측 모델	커머스 TM 고객 중 Call Time 정보가 정확한 MD상품의 Call 대상으로 한 예측모델 ※ TM : TeleMarketing
Call time 예측 모델	Call 성공 예측 모델을 기반으로 시간대별 성공률을 비교/분석하여 AM/PM으로 구분하여 개발
재구매 판매 예측 모델	최근 1년 이내 한 번 이상 구매 경험이 있는 고객을 대상으로 한 예측모델

기대효과

개발된 예측 모델을 기반으로 마케팅 시나리오 수립을 통해 커머스 마케팅 업무 효율성 극대화

07 | 미래에셋생명

유지율 예측 시스템 구축

프로젝트 기간

사업기간 : 3개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 보험시장이 성숙기로 들어서면서 보험 유지율에 대한 관심이 증가하고 계약 단계에서부터 고객관리의 필요성 증가
- 금융감독원의 불완전판매 감축방안에 적용하기 위해 빅데이터를 활용하여 불완전판매계약을 사전에 제거할 수 있는 시스템 도입의 필요성 대두

분석내용

유지율 예측 모델

계약, 고객, 설계사, 접촉 속성 정보 중 유지율에 영향을 미칠 수 있는 항목과 유의성 검정을 진행하고 최종 선정된 항목을 기반으로 FC/TFC/AM 채널에 대한 13/25회차 유지율 예측 모델을 개발함 (스코어카드 방법론 적용)
※ FC : Financial Consultant(설계사)
※ TFC : TeleMarketing 지점
※ AM : 대리점

기대효과

분석을 통한 새로운 인사이트를 발굴하고 고객과의 체계적인 계약관리로 완전판매율과 영업이익 향상 및 보험계약 건전성을 강화하고자 함

08 | 동양생명

유지율 예측모델 개발

프로젝트 기간

사업기간 : 1개월

활용 Tool

SAS

배경 및 목적

- 보험계약 이후에 보험유지율 향상을 위한 고객관리 효과에 한계가 있음
- 동종 업계(신한/미래에셋/ING/메트라이프 등)의 신계약 유지율 예측모델 관심도 증가에 따라 경쟁력 강화 필요

분석내용

**신계약 유지확률
예측 모델**

계약, 고객, 설계사, 점포 속성의 데이터를 기반으로 유지율에 영향을 미칠 수 있는 항목과 유의성 검정을 진행하고, 유의항목 구간화를 통해 GA/FC&DM/DM 채널의 각 7/13/25회차 예측모델을 개발함

- ※ GA : General Agency
- ※ FC : Financial Consultant(설계사)
- ※ DM : Direct Mall

기대효과

FC의 유지율 인식 제고 및 불건전 계약 유입 차단 등 신계약 품질 향상 및 영업 효율 선진화를 위한 새로운 개념의 유지율 분석 시스템 도입

09 | ING생명

민원 예측모델 및 FC 이탈 예측모델 개발

프로젝트 기간

사업기간 : 2.5개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 외부에서 민원을 제기할 고객을 사전에 예측하여 민원을 방지하고, FC의 이탈 시점을 예측하여 손실 방지 목적
음성, 상담 등 비정형 데이터를 내부 정형데이터와 연계하여 Insight를 도출하고, 향후 활용방안 수립 필요
- ※ FC : Financial Consultant

분석내용

민원 예측 모델

- 가설 기반의 정형/비정형 데이터 분석을 통해 대외민원에 영향을 주는 변수들을 도출
(고객/행동/상품/FC별)
- 로지스틱 회귀분석을 통해 대외민원 발생을 예측하는 모델 수립 및 정확도 검증

FC 이탈 예측 모델

- 인터뷰 등 현황분석에 기반하여 FC의 기본속성, 고객정보, 채용정보 등 활동 및 성과 관련
가설을 수립함
- 의사결정나무 알고리즘을 활용하여 FC 세분화 및 Profiling 결과에 따른 모델링을 진행함

기대효과

FC의 유지율 인식 제고 및 불건전 계약 유입 차단 등 신계약 품질 향상 및 영업 효율 선진화를 위한
새로운 개념의 유지율 분석 시스템 도입

10 | 신한생명

신계약 예측 유지율 시스템

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 설계사 유지율 인식제고 및 불건전 계약 유입차단 등 신계약 품질의 질적 향상 및 영업효율 선진화를 위한 새로운 개념의 유지율 분석 시스템 필요
- 설계사가 신계약의 계약정보 입력 시, 해당 계약이 7/13/25회차까지 유지될 수 있는 확률을 조회하는 시스템 구축

분석내용

유지율 예측
스코어링 모델

- 신계약 정보 입력 시점에 설계사가 해당 계약의 회차별 유지확률 조회를 통해 계약관리에 참고
- 설계사, 계약, 고객 속성정보에서 유의미한 변수 추출하여 스코어링 모델 개발
향후, 상품 및 채널, 추가변수 활용한 고도화 필요
- 로지스틱 회귀분석을 활용하여 7/13/25회차 유지율 모델 개발

기대효과

설계사의 유지율 인식 제고 및 선진화된 계약관리를 통해 이익 상승

11 | 교보생명

우수 FP 리크루팅 예측모델 개발

프로젝트 기간

사업기간 : 4개월

활용 Tool

IBM SPSS Modeler 15.0

배경 및 목적

- 개인정보보호 강화 등 금융환경의 변화와 FP 채용시장 경쟁의 심화로 인한 후보발굴의 어려움 존재
 - 최근 위촉된 신인FP의 약 47.6%가 기계약자임을 고려할 때 기계약자로부터의 우수 FP 발굴 노력이 필요함
- ※ FP : Financial Planner

분석내용
**FP 활동가능성
예측모델**

- FP(보험설계사)로 위촉된 고객(기계약자)의 특성을 분석하여 FP로 채용될 가능성을 예측하는 모델 생성
- 데이터 분석을 기반으로 한 마이닝 모델과 비즈니스 경험기반의 휴리스틱 모델을 조합하여 모델 개발

**우수 FP후보
예측모델**

- FP로 위촉 후에 우수 FP로 성장할 가능성이 높은 기계약자를 예측하는 모델 개발

**FP 리크루팅
프로세스 진단**

- 운영중인 FP채용 프로세스에 대한 이슈 및 개선 의견을 파악을 통한 개선과제 및 예측모델 운영 방안 도출

기대효과

심화된 시장경쟁에서 주도권 선취 및 우수 FP발굴로 영업력 상승

12 | 알리안츠생명

AA 조직 목표 관리 및 평가 통계 컨설팅

프로젝트 기간

사업기간 : 1개월

활용 Tool

IBM SPSS Modeler 15.0
Client/Server

배경 및 목적

- 수작업으로 진행중인 영업채널 별 고객 수 현황 파악 개선 및 모니터링 필요
- 영업 채널 별 효율적인 고객관리를 위해 목표 고객 수의 주기적인 조정 및 인사 이동자의 평가 프로세스 자동화

분석내용

AA조직별 목표 조정

- IBM SPSS Modeler를 활용하여 수작업으로 진행되던 실적/변동 고객 수 산출 작업시간 단축
- 엑셀을 이용하여 진행되던 목표 조정 및 달성을 산출 작업을 반자동화하여 Human error를 제외함으로써 정확도 향상
- 이상치 지점 모니터링을 위해 다양한 이상치 지점 탐지 방법을 개발하여 효율적인 지점 관리

인사 이동자 평가

- 부서 간 인사 이동자에 대한 합리적인 평가 방안 개발
- 정확한 인사 정보 수집이 불가하여, 피 평가자가 직접 근무 이력을 기입하고, 이에 대한 직원 평가 점수가 산출될 수 있도록 엑셀 함수 개발

기대효과

심화된 시장경쟁에서 주도권 선취 및 우수 FP발굴로 영업력 상승

12 | 알리안츠생명

마이닝 모델 업그레이드 및 추가 모델 개발

프로젝트 기간

사업기간 : 3개월

활용 Tool

IBM SPSS Modeler 14.2
Client/Server

배경 및 목적

새로운 채널/상품 마케팅 모델 개발 및 환경 구축

분석내용

이탈방지 모델 업그레이드

- 정상유지 고객을 대상으로 이탈(실효 및 해약)할 가능성이 높은 고객을 선별하는 모델
- 최근 데이터 반영, 추가 변수 사용, 스코어카드 모델을 혼합한 모델 사용

추가가입 모델 업그레이드

- Up Selling : 계약자 별 추가 가입 가능성 점수와 추천상품을 예측하는 모델
- 최근 데이터 반영, 추가 변수 사용
- 중복상품 추천을 막기 위해 순차 연관성 모델을 제외한 일반 연관성 모델만 사용

Win-Back 모델

- 이탈고객 중 DM 채널을 통해 Win-Back 가능성이 높은 고객을 예측하는 모델 개발
- 마이닝 스코어를 활용한 Win-Back 대상자 추출의 자동화

신상품 추가가입 모델

- 신상품 출시에 신속하고 유연성 있게 타겟팅 하기 위해 신상품 마트와 상품별 스코어카드 모델 개발

기대효과

모델 고도화 및 신모델 구축으로 새로운 채널과 상품 마케팅 효과 상승 및 효율적인 관리 가능

12 | 알리안츠생명

CRM
모델 개발

프로젝트 기간

1차 사업기간 : 3개월

2차 사업기간 : 2개월

활용 Tool

Clementine 12.0
Client/Server

배경 및 목적

현재 계약자 중심의 고객관리 모델 외 Advisor로의 분석 영역 진화가 필요하고 보험 판매 채널 증가하는 등 고객 관리 체계의 확장 및 데이터 마이닝 시스템의 활용도를 증진시키기 위한 노력 필요

분석내용

신규 가입 모델

- POS의 가망고객, 계약 관련자들을 대상으로 신규가입점수 산출 모델 개발

이탈 방지 모델

- 정상유지 고객을 대상으로 실효될 가능성이 높은 고객을 선별하여 모델 개발

추가가입 모델

- Up Selling : 계약자 별 추가 가입 가능성 점수와 추천 상품을 예측 모델 개발
- Cross Selling : DM 대상 고객 기반으로 추가가입가능 점수 산출 모델 개발

어드바이저
리크루팅 모델

- 고객의 개인 특성 및 계약정보 등을 이용하여 어드바이저로 전환 가능성이 높은 고객 선별 모델

기대효과

CRM 모델을 기반으로 시스템 활용도를 높여, 고객관리 프로세스 및 효율증대

13 | NH농협중앙회

모바일 빅데이터
구독서비스 개발

프로젝트 기간

사업기간 : 6개월

활용 Tool

HIVE & Impala SQL,
Cloudera Data Science
Workbench(CDSW)

배경 및 목적

N-hub(농협 내 정보제공 서비스포털) 데이터를 이용해 범농협 임직원을 대상으로 농산물 종합정보, 축산물 종합정보, 하나로마트 매출 관련 컨텐츠를 만들고 임직원이 개별적으로 구독할 수 있도록 주기적으로 컨텐츠 제공

분석내용

농산물 종합정보

- 시장별 당일 경매 결과, 지역별 작물 출하정보, 지역별 작물 소매가격 정보 등 제공

축산물 종합정보

- 우시장 주기별 리포트, 우시장 가격 추세 등 데이터 정합성 확인 및 컨텐츠 형태로 제공

하나로마트
매출돌보기

- 카드 데이터를 이용한 상권 현황 정보 제공
- POS매출 현황 정보로 매장별 상품판매 동향 확인 서비스 제공

기대효과

임직원의 전문성 강화 및 농축산물 트렌드 정보를 통해 효율적인 업무 운영 제고

14 | NH농협중앙회

농산물 가격 예측 모델 성능 개선 및 신규 작물 가격예측모델 추가 개발

프로젝트 기간

사업기간 : 6개월

활용 Tool

Python, HIVE & Impala SQL
 Cloudera Data Science
 Workbench(CDSW)

배경 및 목적

- 수급안정이 필요한 10대 농산물의 가격예측 서비스 제공
- 기존 6개 작물(배추, 감자, 대파, 무, 사과, 풋고추) 가격 예측 모델 성능 개선
- 신규 4개 작물(마늘, 양파, 토마토, 배) 가격 예측 모델 추가 개발

분석내용

농산물 가격
예측 모델링

- 주요 작물 별로 향후 4주간의 주간 평균경락가격(원/kg)을 예측하는 모델을 개발하고, 예측 가격 상하한을 제공함
 - 기존에 전처리부터 예측값 산출까지의 일원화로 제공되던 프로세스를 각각의 모듈로 분리하는 방향으로 개선해 유연한 모델 운영 가능하도록 개선
 - 변동계수가 큰 일부 작물(사과, 풋고추)을 제외한 8개 작물에 대해 MAPE 15% 내 성능 달성

농산물 가격에
영향을 미치는
주요인자 발굴

- Boruta 알고리즘을 이용해 농산물 별로 가격에 영향을 미치는 변수를 도출하고 마트에 적재
- 4주 주기로 주요인자를 자동으로 도출하는 프로세스 구축

기대효과

수급안정이 필요한 10대 농산물의 가격예측 서비스를 제공해 가격 변동성을 사전에 예측해 효과적으로 의사결정을 내릴 수 있도록 지원하고 시장 안정화와 생산, 유통 효율성 향상

15 | NH농협중앙회

출장우 유사도 분석 모델 개발

프로젝트 기간

사업기간 : 6개월

활용 Tool

Python, HIVE & Impala SQL
Cloudera Data Science
Workbench(CDSW)

배경 및 목적

N-hub(농협 내 정보제공 서비스포털) 축산물 종합정보에 임직원 대상으로 제공할 출장우 유사도 분석 모델을 개발하여 일 단위로 유사도 분석 정보 및 가격 추세 예측 서비스를 함께 제공

분석내용

유사도 분석 모델

- 내부 한우종합, 한우핵심 DB 출장우 내역 및 외부 축평원 데이터를 기반하여 유클리디안 거리로 유사도 분석 모델 개발
- 가축시장 가격예측으로 전체 시장 동향을 예측

기대효과

유사한 출장우 개별 개체 뿐만 아니라 전체 동향을 함께 살펴볼 수 있기에 사전에 출장우를 판매하기 전 가격을 효율적으로 예측하여 의사결정을 도움

국내 프로젝트 사례집

물류/유통 (Logistics / Merchandising & Distribution)

- | | | | |
|----|--------|----|--------|
| 01 | 스타벅스 | 06 | CJ대한통운 |
| 02 | 삼성SDS | 07 | 신세계백화점 |
| 03 | 신세계면세점 | 08 | 신세계몰 |
| 04 | 한살림 | 09 | 현대상선 |
| 05 | 휴마스온 | | |

01 | 스타벅스

원부자재 수요예측 모형 개발

프로젝트 기간

사업기간 : 4개월

활용 Tool

Databricks, Python,
sparkSQL

배경 및 목적

- 기존 운영중인 권고발주량의 활용성 개선 및 최적 재고관리 필요성 증대
- 음료 판매량 수요예측 모델 개발 및 고도화를 통한 원부자재 사용량 수요예측 정확도 개선을 목적으로 함

분석내용

매장 세분화

- 매장에서 발생하는 판매량 기반의 정량적 특성에 기반하여 매장 세분화
 - 판매 유형, 안정성, 캘린더, 기상 등 26개 영역의 세분화 지표(등급) 개발
 - 지표를 통한 유사 매장 분류 및 트렌드 등 매장별 최적화 된 관리 가능
- 매장 세분화 결과의 안정성 평가를 위해 직전 세분화 등급에 대한 PSI 지수 산출
 - 편차 PSI : 세분화 별 등급 내 속한 매장의 변동성을 평가하여 매장의 변화 및 판매 변동 확인
 - 매장 수 PSI : 세분화 등급별 매장 수의 분포 변화를 평가하여 세분화 등급의 안정성 확인

수요예측 모델 개발

- 매장별 음료별 판매 특성에 따라 ML 및 RULE 예측 대상을 구분
 - 판매량이 낮거나 변동성이 높은 경우 패턴 학습이 어려워 RULE을 통한 예측 수행
- ML모델 학습 시 유사한 매장X음료간 학습을 위해 유사한 판매 특성을 가진 매장별 학습 수행
 - 매장 세분화 지표별 학습 : 판매량 패턴 및 변동성이 안정적인 경우
 - 매장 상권별 학습 : 월별 판매량 변동이 불규칙하여 지역 및 상권에 대한 특성을 띠는 경우
 - 개별 매장별 학습 : 개별적인 특성을 갖고 있는 경우
- 신규 매장 및 음료 출시 시 초기예측을 통한 Cold Start를 방지하기 위한 방안 수립

기대효과

권고발주량 최적화를 통한 재고 관리 비용 감소 및 매장 점주의 신뢰도 상승

02 | 삼성SDS

물류 Analytics 역량강화 분석 시스템 개발

프로젝트 기간

3차 사업기간 : 8개월

4차 사업기간 : 8개월

활용 Tool

Brightics ML, Python,
Vensim

배경 및 목적

- 전자 기준 최적 ETA 예측 모델 개발을 통한 예측 정확도 제고 및 Brightics 기반 프로세스 자동화
- Risk 센싱 모델, 희망봉 우회로직, Port Congestion로직 개발을 통한 글로벌 이슈 대응 및 ETA 예측 모델 고도화
- 통계&ML 기반의 선적항 ETD 예측 모델을 통해 출항 지연 관리 및 효율적인 물동량 관리 지원
- 기존인자 검증 및 신규인자/파생인자 생성을 통한 모델 고도화, 항공 운임 시황 모델 재구축(원가기반 모델)

분석내용

전자 기준 최적 ETA 산출	<ul style="list-style-type: none"> 가설 기반 지연 데이터 탐색을 통해 지연/조기도착 특성정보를 탐색하고 모델 고도화 및 검증 <ul style="list-style-type: none"> - 북미/구주/아프리카 권역에서 남은 예상 일수가 7일 이하이면서 ETA가 일요일인 경우 지연 발생 비율이 높음 - 전체 권역에서 ETA가 남은 예상 일수가 7일 초과이면서 ETA가 일요일인 경우 지연 발생 비율 높음 - 환적 대상 선박 중 스케줄 상의 예상 지연 일수가 14일 이상인 경우 지연 발생 비율 높음 ETA Coverage 100% 달성 <ul style="list-style-type: none"> - 스케줄이 없는 경우 기존 전자 제공 Port ETA 산출로직을 통해 ETA 제공
글로벌 이슈 대응 로직개발	<ul style="list-style-type: none"> Risk센싱 모델 개발 <ul style="list-style-type: none"> - 일부 권역 및 Port에서 Port Skip 또는 지연으로 인한 Risk를 사전에 탐지하는 모델 이상 상황 발생 대응 로직 개발(ETA 예측 모델 고도화) <ul style="list-style-type: none"> - 희망봉 우회 로직: 수에즈를 통과하는 선박이 희망봉을 우회하면서 발생하는 지연정보 반영 - Port Congestion 로직: 외항영역 대기 선박 수 증가로 인한 체류시간 지연정보 반영
선적항 ETD 예측	<ul style="list-style-type: none"> FIS 데이터&스케줄 데이터 분석을 통해 신규 모델 개발완료 및 ETD예측 Coverage 80% 달성 통계모델&ML모델 기반의 하이브리드 모델 구축 <ul style="list-style-type: none"> - 분석 마트 예측에 필요한 주요 인자를 도출하여 마트를 적재하고 모델 자동 학습 및 예측 값 산출 프로세스 구축
운임시황 모델 고도화	<ul style="list-style-type: none"> 기존 인자 검증 및 보완, 신규 인자 발굴과 같은 입력 인자 관리를 통해 예측 모델 고도화 ‘수요-공급’ 기반 항공 운임시황 예측 모델을 사회적/경제적 이슈에 의한 변동성을 고려할 수 있는 ‘원가’기반 예측 모델로 재구축 인자 생성, 데이터 전처리, Report작성 업무 자동화를 통한 효율적인 프로세스 관리

02 | 삼성SDS

물류 Analytics 역량강화 분석 시스템 개발

프로젝트 기간

1차 사업기간 : 8개월

2차 사업기간 : 8개월

활용 Tool

Brightics ML, Python,
Vensim

배경 및 목적

- 선박 화물 ETA 예측 모델과 해상 항공 운임시황 예측 모델은 새로운 데이터를 통해 중요 인자를 추가하고 다양한 알고리즘을 활용하여 기존 모델을 고도화 시키는 것을 목적으로 함
- 운송 실적 및 내륙운송 견적 데이터를 사용하여 신규 잠재 고객에 대한 내륙운송 Backhaul 기회 발굴

분석내용

선박 화물 ETA 예측	<ul style="list-style-type: none"> · POD(도착지) 기준 ETA는 LANE(Port간 이동) 소요시간과 Port 체류시간을 합산하여 산출 <ul style="list-style-type: none"> - LANE 소요시간 예측 모델 <ul style="list-style-type: none"> * 정의 : Port와 Port를 이동하는데 소요되는 시간을 예측(Port 외항 영역을 기준으로 산정) * 내용 : 과거 선사 스케줄 데이터와 실시간 위치 정보(AIS) 데이터를 기반으로 지난 모든 선박의 항적 마스터 정보를 구축. 예측 대상 선박의 이동경로 정보를 기반으로 과거 유사한 항적을 산출하여 예상 소요시간을 추출. - PORT 체류시간 예측 모델 <ul style="list-style-type: none"> * 정의 : Port 접안 영역까지의 이동 시간과 선/하적 시간 등을 합산한 시간을 예측 * 내용 : 과거 실시간 위치 정보(AIS) 데이터를 기반으로 Port의 외항 영역 내 체류 마스터 정보를 구축. 추가로 Port Congestion(혼잡도)을 기반으로 가중치를 적용하여 최종 Port 체류시간 예상 시간 추출. · 선박 화물 ETA 산출 시 예상 운항 경로 및 지역 정보 등을 제공 · Geo-fence 기준 정보 재정비를 통한 ETA/실적 센싱 모델 정확도 향상 · ETA 모델 고도화를 통한 정확도 향상 (1차년도: 70% → 2차년도: 85% 달성)
해상 항공 운임시황 예측	<ul style="list-style-type: none"> · PMI, 생산지수 및 과거 수요 정보 등을 기반으로 수요와 공급을 예측하여 최종적으로 해상 및 항공에 대한 운임을 예측.
내륙운송 Backhaul 기회분석	<ul style="list-style-type: none"> · 출발한 운송이 공차인 상태로 출발지로 되돌아 오지 않도록 도착지 주변 실적을 찾아 다시 출발지 근처로 운송할 수 있게 함. <ul style="list-style-type: none"> * 방안 : Cello운송실적 정보와 내륙운송 실적을 기반으로 Backhaul이 가능한 데이터를 매칭 → 출발지 기준과 도착지 기준으로 Backhaul의 반경, 차량 탑입, 출도착 시간 등을 사용자가 지정하여 매칭

03 | 신세계면세점

디지털 혁신 (DI)을 위한 분석모델 개발

프로젝트 기간

사업기간 : 6개월

활용 Tool

Python, SK Accuinsight+

배경 및 목적

- 고객 관리 및 캠페인 활동, 채널별 추천 상품, 발주업무의 디지털 혁신을 위한 AI 분석모델을 개발하여 각 업무의 효율화를 제고를 목표로 함

분석내용

스코어링 모델

- 고객가치, 구매, 행동관점의 선호도라는 기준을 생성하고, 선호도별 고객의 점수 및 등급을 산출하여 타겟 마케팅에 활용할 수 있도록 개발함
- 42개 항목의 약 2000여개 선호도 생성

상품추천

- 연관상품 관점(구매기반)과 유사상품 관점(웹 로그 기반)의 Item to Item 추천적용
- 연령과 성별로 세분화를 구성하여 연관상품과 유사상품 모델 개발(Item2Vec 알고리즘)
- 고객이 식별되지 않는 경우를 위해 통합모델을 구성
- 상품 특성을 고려하여 필터링 조건을 적용
- 당사에 최적화된 추천 로직 보유 및 관리할 수 있는 체계 마련

수요예측

- 국산/외산 상품의 발주 주기(주&월)를 고려하여 SKU/지점 단위로 수요예측 모델 개발
- 할인행사 정보를 고려한 이상치 설명 및 영향요소 고려
- 단순 통계량(3개월 평균 등)과 혼합하여 정확도 최적화 적용
- 발주량 설정 시 보다 정확한 예상판매량 제공

※ SKU (Stock Keeping Unit) : 상품의 재고관리를 위한 최소 분류 단위. 개별적인 상품에 대해 재고관리 목적으로 추적이 용이하도록 하기 위해 사용되는 식별관리코드

기대효과

분석모델을 기반으로 고객 및 상품관리 강화와 타겟 마케팅을 통한 영업이익 상승, 사전 수요예측을 통한 비용감소

04 | 한살림

한살림

물품 수요예측 모델 개발

프로젝트 기간

1차 사업기간 : 1개월

2차 사업기간 : 5개월

활용 Tool

IBM SPSS Modeler, Python

배경 및 목적

- 물품 판매량을 예측하여 고객 주문 발생 전 공급처에 미리 발주를 넣는 ‘선발주 시스템’을 통해 배송시간 단축
- 한살림 발주 프로세스를 고려한 매장 별 물품 수요예측을 통한 AI 알고리즘 기반 재고 관리 시스템 구축

분석내용마트상품 판매량
예측 모델

- 과거 판매정보(판매량, 판매시점, 쿠폰/할인 행사, 구매고객, 대체재)와 캘린더/날씨 정보를 활용하여 ‘상품별 판매 패턴’을 파악
- 상품별 일별/주별/월별 판매량 예측 모델 개발

매장 별 물품
수요예측 모델

- 매장 및 물품의 판매정보 특성 및 패턴을 고려한 맞춤형 모델 개발
- 행사정보, 캘린더, 날씨 등 다양한 외부변수 및 파생변수 활용
- 기대손실 적용 및 결품 보정으로 예측 정확도 확보

기대효과

- ‘선발주 시스템’을 도입하여 배송시간을 단축시키고 ‘새벽 배송’ 또는 ‘신선식품 배송’ 등 차별화된 고객서비스 제공
- 결품 및 과잉재고를 고려하여 손실을 최소화하고 최적 판매량을 예측하여 효율적인 재고 관리 실시

05 | 휴머스온

온라인 쇼핑몰 개인화 상품 추천 모델

프로젝트 기간

사업기간 : 3개월

활용 Tool

Python

배경 및 목적

- 여러 쇼핑몰이 통합되어 있는 휴머스온은 쇼핑몰 별 상품명 및 카테고리 체계가 상이하여 제한적인 추천만 가능. 통합 카테고리 기반의 개인화 상품추천 모델을 개발하여 다양한 상품추천을 목표함

분석내용

온라인 쇼핑몰
개인화 상품추천 모델

- 다수 쇼핑몰을 하나의 통합 카테고리 체계로 구축하였으며, 통합 카테고리 기반의 상품추천 로직 개발진행
- 고객의 구매욕구 해소 여부를 기준으로 분석에 필요한 타겟 고객군 정의
- 구매욕구 해결 고객군 / 구매욕구 미해소 & 관심상품 등록 고객군 / 비활동 및 신규 고객군 → 3개 고객군으로 개발
- 연관성분석 및 콘텐츠 기반 유사도 모델 적용

기대효과

- 고객의 니즈에 맞는 상품 추천으로 고객 구매력 상승 및 마케팅 효율화 증대를 기대함
- 실제 상품 추천 테스트 결과 기준에 진행하던 추천 적중률 대비 개선된 적중률을 보임
- 통합 카테고리 적용으로 인한 신규 쇼핑몰 고객 유입 시 타 쇼핑몰 이용 고객의 분석 결과를 적용할 수 있는 확장성 확보

06 | CJ대한통운

택배 물동량 예측 시스템 구축 및 고도화

프로젝트 기간

1차 사업기간 : 4개월

2차 사업기간 : 4개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

택배 이용 고객의 증가로 매일 5백만 건 이상의 운송장 정보가 축적됨에 따라 분석모델 발굴 및 구축 필요성 증대

분석내용
**물동량 수요
예측 모델**

운송장/캘린더/기준점소/고객/실적/지역/날씨 등의 정보
외부 홈쇼핑 정보 및 트렌드(추세) 변화를 반영

342개의 SUB 터미널에 대해 월간, 주간, 당일 물동량 예측 선형회귀 모델 개발

기대효과

- 통합적 분석을 토대로, Sub별 특성을 반영한 효율적이고 정확한 배차 가능
- 영업 경쟁력 강화 및 자원 운영의 효율화 증대



07 | 신세계백화점

백화점 개인화 캠페인 구축

프로젝트 기간

사업기간 : 6개월

활용 Tool

R

배경 및 목적

- 고객의 니즈와 상품의 다양성이 증가하면서 개인의 요구에 맞는 추천 서비스를 통해 고객 편의 제공 및 구매 전환율 제고를 위한 필요성 증대

분석내용

개인화
추천서비스 모델

- 고객 가치(RFM지수), 서비스 이용지수, PC 다양성을 축으로 고객을 세분화하고, 구매 MD를 중심으로 상세 세분화를 진행함
 - 고객정보/MD별 구매/선호도/고객가치/활동 정보를 기반으로 브랜드별 추천점수에 대한 예측 모델을 개발함(선형모델, 부스팅, 워드투백 방법론 적용)
- ※ RFM지수 : 규모/빈도/최근성 기준의 가치 있는 고객을 추출하여 고객을 분류하는 방법

기대효과

- 고객의 구매활동을 촉진하여 구매 욕구와 만족도를 증진하여 지속적 구매 유도 역할 수행
- 개인화 된 서비스를 통해 신세계 백화점에 대한 고객의 로열티 증가



08 | 신세계몰

CRM 모델

Migration 및

고객 세분화

프로젝트 기간

사업기간 : 1개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 고객의 구매패턴 및 상품 연관성에 근거한 상품 추천, 세분화된 고객 그룹 별 맞춤형 캠페인 전략 수립 등 최적화된 CRM분석을 통한 차별화된 고객 서비스 실현을 위한 고객 및 마케팅 시스템 구축

분석내용

CRM 마이닝 모델
Migration

- 기존에 운영 중인 5개의 CRM 모델에 대해 파악하고 CRM 모델 Migration 시 추가 모델 개발 사항 정리

충성도 기반
고객세분화

- 기존에 활용하고 있는 고객 세분화 모델은 업무 담당자가 직관적으로 이해하기 어려운 모델로 개발되어 있음. 이에 효율성, 활용성, 타당성, 이해력 등을 기준으로 고객세분화 모델을 재개발 함
- 고객 충성도를 기준으로 고객을 세분화 하는 방법론 개발
- 충성도 산출 방법으로 구매력, 유지력, 다양성, 관심도, 파급성의 변수를 활용하여 모델 개발 접근
- 최종적으로 구매력(총 구매 금액), 유지력(경과 기간, 구매 월 빈도) 두 개 척도를 이용하여 고객을 세분화하는 방법론 제안

기대효과

- 고객 가치, 특성을 반영한 고객 세분화 체계를 수립하여 고객 그룹별 맞춤 전략을 실현하고 고객의 구매력 상승 및 만족도 증진
- 고객 및 마케팅 분석 시스템을 통한 고객 관리 향상

09 | 현대상선

기기수급 예측 시스템 구축

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- Empty Container의 수요/공급 예측 정확도 향상을 위한 기기 수요 예측 시스템 구축

분석내용

Empty
Container
수급 예측

- 총 8,896 개 (전체 4,448개 Location, Type, Size x Export Release, Import Return)의 DFU에 대하여 3주 / 6주 이동평균법을 이용하여 수요 예측
- 전체 DFU 중 주간 평균 수량이 10 이상인 DFU 선택하여 SPSS 자동 선택 통계 모델 (Expert Modeler) 기법을 이용하여 총 627개(Export ; 285개, Import ; 342개)의 DFU에 대하여 수요 예측
- 자동 모델을 활용한 오차율이 높고 비즈니스 중요도가 높은 총 35개의 DFU에 대해 데이터 상세 탐색을 동반한 정교 모델(이상치 보정이나 모델 적용 기간 수정 등의 작업)을 개발함

※ DFU(Demand Forecasting Unit) : 예측의 기본 단위.
Location x Container Type/Size를 의미함.

기대효과

기기수급 예측 시스템 구현을 통해 오차 발생 부분에 대한 원인 및 대책을 수립함으로써 업무의 효율성 및 수익 극대화



국내 프로젝트 사례집

엔터테인먼트 (Entertainment)

01 두산베어스

02 파라다이스 세가사미

03 에버랜드

01 | 두산베어스

연봉고과
시스템 구축

프로젝트 기간

사업기간 : 5개월

활용 Tool

Python, MySQL



배경 및 목적

- 수동 연봉 책정으로 주관적 평가에 의존하여 평가 항목과 배점 기준에서 객관성 부족 문제 발생
- 두산베어스에 최적화된 자동 고과 산정 프로세스 및 연봉 평가 시스템 구축 필요

연봉고과 시스템 구축

신규 고과 항목 선정	<ul style="list-style-type: none"> 데이터 관점의 검증 및 탐색으로 기존 고과 항목 선별 비즈니스 실무 중심의 카테고리 정의 및 시뮬레이션을 통한 배점 최적화
고과 점수 산출 프로세스 개발	<ul style="list-style-type: none"> 전력 분석 웹 포탈 DB 및 수집 데이터를 활용한 경기 별 고과 기록 산출 경기 실적 기반의 다양한 상황에 따른 가중치를 적용한 점수 집계
차기 연봉 산출 프로세스 개발	<ul style="list-style-type: none"> 수집된 데이터 중심의 연봉 마트 구축 및 고과 점수 모델 적용 과거 연봉 인상률을 기반으로 차기 연봉 및 고과 등급표 생성

기대효과

- 데이터 기반의 연봉 산출 자동화를 통한 타당성 확보 및 선수와 구단 간 합리적인 연봉 체결을 통한 비용 절감 효과
- 샐러리캡을 포함하여 다양한 상황을 고려한 시뮬레이션 수행 및 결과 제공

02 | 파라다이스 세가사미

카지노 방문 고객수 예측 모델 개발

프로젝트 기간

1차 사업기간 : 2개월

2차 사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 카지노 딜러의 적정 배치를 통한 인력 운영비용 절감 목적.
- 딜러의 수는 운영되는 게임 테이블 수에 따라 정해지므로, 카지노 방문고객 및 운영 테이블 수를 예측필요.

분석내용

방문 고객수
예측모델 개발

- 고객관점 분석대상 : VIP, MASS(일반) 고객
- 게임관점 분석대상 : BJ(블랙잭) 등 총 6개 게임종목
- 고객 그룹별(Seg), 게임 종목별, 영업시간별(1/2/3부), 영업일별 모델 적용
- 선형 회귀분석을 통한 예측모델 개발

운영 테이블 수
산출 로직 수립

- 게임 종목별 영업시간/시간대에 방문하는 고객수 대비 테이블 수 산출 및 적용
- 게임 종목별 일/시간대별(00시 ~ 23시)

기대효과

- 예측 결과값을 기반으로 사전에 딜러 배치 계획을 수립할 수 있음.
- 적정 수준의 인력배치를 통한 비용 절감 효과를 기대할 수 있음.



03 | 에버랜드

입장객 수 예측모델 개발

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 기존 운영하던 입장객 수 예측모델의 오차율 증가로 수작업을 통해 예측하였지만, 정확도 및 업무 효율성 저하.
- 입장객 수 예측업무 프로세스를 효율적으로 정립하고 모델기반의 예측력 상승 필요.

분석내용

분석대상 및 활용변수

- 회원여부(회원/개인)와 주중(평일/휴일)으로 모델 적용
- 캘린더, 기후, 축제여부, 과거 입장데이터, 프로모션 정보 등 예측 영향요소로 활용

장기 입장객 예측
(Baseline 예측모델)

- 날씨가 정상 기후인 경우로만 예측하여, 향후 1년간 장기 입장객 목표설정을 위한 장기 입장객 예측모델 개발.

단기 입장객 예측

- 모든 기후(맑음, 눈, 비 등)를 고려하여 신경망 및 의사결정나무 모델을 활용한 단기 입장객 예측모델 개발.
- 신경망과 의사결정나무 모델을 양상불하여 결과 활용.

기대효과

입장객 수 예측모델을 활용하여, 효율적인 인력 배치와
사전 준비로 비용절감 및 이익상승 기대



03 | 에버랜드

구매상품 간
연관성 분석

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 고객이 구매하는 상품 간의 연관성 분석 및 방문하는 매장의 순차분석을 통한 마케팅 활성화로 신규 비즈니스 Insight를 발굴하는 것이 목적.

분석내용

분석대상 정의	<ul style="list-style-type: none"> · 동반하여 입장하는 개인고객의 유형을 비즈니스적 규칙을 적용하여 정의. <ul style="list-style-type: none"> - 미취학동반, 소인동반, 청소년동반, 청소년Only, 연인친구 등 · 동반유형별로 매장/상품 간 패턴 분석하여 전략마케팅 기획 및 MD판매 결과 검증 활용
매장 간 순차구매 패턴 분석	<ul style="list-style-type: none"> · 고객이 방문하는 매장 간의 연계특성을 고려하여 특정 매장의 활성화 및 매출 증대 · 정의된 동반유형별 방문 매장특성을 활용한 프로모션 기획 가능
상품 간 구매 연관성 분석	<ul style="list-style-type: none"> · 상품 간의 연관성 분석을 통해 자주 판매되는 상품과, 실적이 미미한 상품을 묶어, 전 상품에 대한 판매증가
이용권 할인율 별 객단가 차이 분석	<ul style="list-style-type: none"> · 이용권 할인율을 기준으로 고객 그룹을 분리하여, 그룹별로 매장과 상품에서 구매하는 단가 차이를 분석하여, 프로모션 기획 가능

기대효과

구매상품 간 연관성 분석을 통해, 상품 판매를 통한 이익증대와 다양한 프로모션으로 매장활성화 가능

국내 프로젝트 사례집

통신 & 인터넷 (Telecommunication & Internet)

01 SK브로드밴드

02 LG U+

03 SK텔레콤

04 롯데정보통신

05 KT

01 | SK브로드밴드

마케팅 스코어링 시스템 구축

프로젝트 기간

사업기간 : 9개월

활용 Tool

R

배경 및 목적

- 마케터가 TM(텔레마케팅)진행 시 효율적인 타겟 고객군 설정 필요
- 고객이탈 방지 및 타 상품 구매 가능성을 예측하기 위해 스코어링 시스템 구축 목적

분석내용

FD Rule 개발	<ul style="list-style-type: none"> · 가입기간이 1~15개월 이내인 이용고객 대상 · 고객이 구매한 상품의 약정기간이 만기하지 않았음에도, 조기에 해지하는 이상거래 패턴감지 규칙 개발
예측 Scoring 개발 (Retention)	<ul style="list-style-type: none"> · 번들(TV&인터넷) 이용고객, TV단독 이용고객 대상 · 분석대상의 서비스 이용기간에 따라 세분화 하여 각 세분화된 고객군 특성에 기반한 예측 모델을 개발
예측 Scoring 개발 (Value-up)	<ul style="list-style-type: none"> · SK브로드밴드에서 제공하는 서비스별로 추가가치를 제공할 수 있는 상품을 정의. 정의된 상품별 고객의 가입 가능성을 예측하는 모델을 개발

기대효과

- 스코어링 시스템을 통한 효율적인 타겟 마케팅으로 기존에 운영하던 인바운드 해지방어 캠페인에 추가로 아웃바운드 해지방어 캠페인 효과가 상승함.
- Value-Up Scoring 모델을 활용하여 고객의 가입 가능성을 예측하여 Cross/Up Sell 타겟 마케팅 가능

02 | LG U+

모바일/홈 해지방어 예측 모델 개발

프로젝트 기간

사업기간 : 4개월

활용 Tool

SAS

배경 및 목적

- 마케터가 TM(텔레마케팅)진행 시 효율적인 타겟 고객군 설정 필요
- 모바일 및 홈(유선인터넷) 가입고객의 이탈 방지 가능성을 예측하는 스코어링 모델 개발 목적

분석내용

모바일 고객 이탈 방지 모델	모바일 서비스를 이용중인 고객의 이탈을 사전에 방지하기 위하여 이탈방지 예측모델 개발
홈(인터넷) 고객 이탈 방지 모델	인터넷/TV 서비스를 이용중인 고객의 이탈을 사전에 방지하기 위한 이탈방지 예측모델 개발

기대효과

스코어링 시스템을 통한 효율적인 타겟 마케팅으로 기존에 운영하던 인바운드 해지방어 캠페인에
추가로 아웃바운드 해지방어 캠페인 효과가 상승함.

03 | SK텔레콤

고객 패턴분석 기반 프로파일 DB 구축

프로젝트 기간

사업기간 : 4개월

활용 Tool

R

배경 및 목적

- 현재 서비스 중인 CBS의 기업고객이 원하는 타겟팅 조건을 반영하기 위한 고객별 세분화 기준 및 지수가 필요함

분석내용

성향 지수

- Biz. 활용 관점에서 소비, 관계, 활동, 충성도, 접촉/반응 중심의 7가지 성향을 산출함 (Life Stage, 가족유형, 구매력, 선호채널, 경제활동, SKT충성도, 활동성향)

관심사 지수

- 고객의 App 사용, 통화, T멤버십 사용, 위치이력 정보, 초콜릿 사용정보를 활용하여 관심사를 식별함. 26개 관심사별 5개의 채널 가중치를 고려하여 관심도 지수 산출

기대효과

- CBS 고객사가 원하는 지역 및 특정 고객에게 맞춤형 푸시 마케팅이 가능함
- 솔루션 경쟁력을 강화시키고 지역별 특성에 대한 추가 분석이 가능함

03 | SK텔레콤

분석모델 통합플랫폼 통계모델

프로젝트 기간

사업기간 : 8개월

활용 Tool

R

배경 및 목적

- 과거 소명 대상은 기존 모델에 의해서만 발견된 건들이기 때문에, 기존에 식별되지 않은 비정상 case를 다른 관점으로 식별할 수 있는 모델의 필요성 대두

분석내용

비정상 탐지 모델 개발

- 스코어카드 모델(로지스틱 회귀분석)과 기계학습 모델(의사결정나무, SVM, 랜덤 포레스트)을 통하여 비정상 case를 적발함
- KNN 및 이상치 탐지 방법을 적용하여 기존의 비정상 case와 유사한 정상 case를 찾아 비정상 건들을 탐지함

기대효과

신규 비정상 탐지 모델은 감사 대상자 중 실제 비정상 적발률을 향상시키고, 적발률을 유지하는 상태에서 비정상 적발 모수를 증가시킬 수 있음

03 | SK텔레콤

SC기지국 혼잡예측

프로젝트 기간

사업기간 : 3개월

활용 Tool

IBM SPSS Modeler 14.1

배경 및 목적

- 기지국별 과다 트래픽 등으로 인한 혼잡지표 생성하여 기지국 관리 필요

분석내용

Smart Controller
시스템 개발

- 기지국 혼잡인지 : Radius 메시지의 사용량을 기준으로 혼잡도 인지 지표개발 및 산출
 - 과다 트래픽 사용자 인지 : 트래픽을 과도하게 사용하는 과다사용자 분류(Heavy-Medium)
 - Reactive QoS 정책사용 : QoS정책을 적용할 사용자 정의 및 분류
- ※ QoS : 서비스품질정책(데이터 흐름 등에 우선순위를 정하여, 데이터 전송의 특정수준의 성능보장)

군집분석을 통한
기지국 분류

- 현실적으로 모든 기지국을 대상으로 개별 스코어모델의 관리가 어려움
- K-Means Clustering 알고리즘을 활용하여, 유사한 기지국 군집 분류 및 검증 작업 진행

기지국 혼잡 인지
스코어 모델 생성

- Radius 데이터를 기반으로 기존 SC모델 개선 및 신규모델 생성
 - 회귀분석을 통해 분류된 기지국 군집별 Congestion Score 모델링 구축
- ※ Radius : Remote Authentication Dial In User Service /
원격 인증 전화 사용자 서비스위치
- ※ SC : 무선통신 접속되어 신호를 수신할 수 있는 기지국
- ※ Congestion Score 모델링 : 혼잡 스코어 모델

기대효과

기지국별 혼잡지표를 활용하여, 트래픽 및 과다 사용자 관리 가능

03 | SK텔레콤

기지국 부하영향 인자분석

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler 14.1

배경 및 목적

- SKT의 LTE 기지국에 부하가 발생함에 따라 부하를 발생시키는 원인과 해당 원인의 영향정도를 탐색

분석내용

LTE 기지국 부하발생 영향요인 탐색 및 영향력 확인

- 기지국 혼잡을 제어할 수 있는 환경을 구축하기 위한 탐색
- 영향요인으로는 고객 수, 세션 수, 머무르는 시간, 접속시도 수 등으로, LTE기지국의 부하여부를 판단하는 장비의 데이터를 결과변수로 활용하여 탐색
- 영향요인의 유의성 검증, 일자별/시간대별/시도표 데이터 패턴 탐색, 영향요인 간의 상관관계 탐색, 로지스틱 회귀분석으로 영향 요인의 영향력 파악을 수행함.

기지국 부하 예측모델 개발

- 수도권 지역의 기지국을 대상으로 부하 빈도가 높은 상위 3개 기지국을 선정하여, 로지스틱 회귀분석, 의사결정 나무모델을 통해 각 기지국별 영향요소 및 영향력과 분석모델의 정확도를 검증함
- 기지국마다 외부환경에 따른 특성이 다르기 때문에, 부하원인 및 영향력에 대한 정확도가 떨어지며, 분석의 필요성에 비해, 매우 높은 정확도를 가진 모델을 개발하기 어려운 한계가 있음.

기대효과

모델 개발의 한계점을 해결할 수 있다면, LTE기지국의 부하를 발생시키는 요인을 대상으로 사전에 관리를 통해 부하발생을 차단할 수 있음

03 | SK텔레콤

App 패턴 분석 시스템 구축

프로젝트 기간

사업기간 : 6개월

활용 Tool

IBM SPSS Modeler 14.1
IBM SPSS Statistics 19.0

배경 및 목적

- 모바일 SNS 서비스의 범위 및 성장으로 인해, 모바일 Traffic이 증가하고 Push성 메시지를 보내는 망에 부하가 발생함
- 트래픽 및 망 부하의 상황에서 신속한 대응으로 고객 서비스 품질 및 업무 효율성 제고 목적

분석내용

Known App 현황 분석	<ul style="list-style-type: none"> 스마트폰의 특정 App에 의해 발생되는 Traffic 데이터를 분석하여, App별 Keep Alive성 Traffic의 발생량을 확인할 수 있음. ※ Keep Alive : 디바이스 간의 데이터 연결이 잘 작동되고 있는지 확인 및 연결 끊김을 방지하기 위한 디바이스 간 주고받는 메시지
Unknown App 현황 분석	<ul style="list-style-type: none"> 스마트폰의 App에서 발생되는 Traffic 중 Keep Alive 특성을 보이는 패턴 및 주기 현황을 분석 이를 통해 특정 IP, PORT에서 발생되는 일정한 크기의 패킷량이 Keep Alive성 Traffic에 해당되는지 여부를 판단할 수 있는 정보를 제공함
Keep Alive 패턴 분석 (Keep Alive 추천 모델)	<ul style="list-style-type: none"> 특정 IP, PORT, 패킷량으로 발생되는 Traffic이 Keep Alive성 Traffic일 확률을 의사결정모델을 통해 도출 Unknown App 현황 분석과 Keep Alive 성 Traffic을 판단할 수 있는 예측 보조 정보로 활용.

기대효과

Keep Alive 및 App Traffic 데이터를 분석하여, 트래픽이 급증 및 망 부하를 유발시키는 영향요소를 파악하고 사전에 방지할 수 있음.

03 | SK텔레콤

서버 사용률 최적화 시스템

프로젝트 기간

사업기간 : 3개월

활용 Tool

IBM SPSS Modeler 14.1

배경 및 목적

- 플랫폼 서비스 인프라에 대한 용량 예측을 통계적 기법을 활용하여 데이터 기반의 정확한 용량 예측정보 제공 목적

분석내용

서버 사용률 최적화
시스템 구축

- 시스템 내 전체적인 추세(Trend)를 파악할 수 있는 예측 정보
- 이상치 자동 보정 및 고객의 의사를 반영할 수 있는 Simulation 수행 기능
- 영향요소 및 각종 이벤트 등을 반영하는 Simulation 예측 정보
- 향후 시스템화 및 자동화가 가능한 예측 정보
- 수집된 Traffic양에 기반한 TPS 정보 제공

※ TPS : 초당 발생하는 트랜잭션 수로, 서비스 성능의 기준 중 한 가지

기대효과

플랫폼 Engineering 업무의 고도화를 지원하는 시스템을 구축할 수 있음

03 | SK텔레콤

과금 검증
체계 고도화

프로젝트 기간

사업기간 : 4개월

활용 Tool

IBM SPSS Modeler 14.1

배경 및 목적

- 스마트폰과 App 열풍으로 데이터 통화량이 급변함에 따라 서버 등 장비 증설비용 및 과금 시스템 장비 관리 Issue 발생
 - 과금 장비의 Traffic에 대한 예측 정보의 생성이 필요
- ※ 과금 시스템 : 선 요금납입 한 고객에 대해, 이용량에 따른 잔액을 감지하는 시스템

분석내용

데이터 과금
검증체계 고도화
예측 모델 개발

- 전체적인 추세(Trend)를 파악할 수 있는 예측 정보 제공
- 이상치의 자동 보정 및 고객의 의사 반영 Simulation 수행 기능
- 영향요소 및 각종 이벤트 등을 반영하는 Simulation 예측 정보 생성
- 향후 시스템화 및 자동화가 가능한 예측 정보 생성
- Traffic 자체 값이 기본적인 계절/요일성 변동추세와 상승/하강 추세를 파악하는 모델 개발

기대효과

정확한 과금 예측모델을 통해 즉각적인 사용자 응대가능 및 추가 이용량으로 발생한 손실 방지

04 | 롯데정보통신

Beacon 동선분석 연구

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- IoT(사물인터넷) 기술의 발달로 수집 가능한 고객 데이터가 증가
- 대형 상업공간에 대한 표준화된 실내 공간 정보에 대한 분석 필요성이 증대

분석내용

고객 동선
분석모델 개발

가상 비콘을 설치하여 얻은 고객의 방문로그 기록을 활용하여, 연관성 분석 및 사회네트워크 분석으로 체류시간, 고객 동선, 허브존에 대한 패턴 및 데이터 인사이트 발굴

※ 허브존 : 고객이 이동한 동선 중 각 동선을 연결하는 연결지점

기대효과

고객 체류시간 및 동선 분석을 기반으로 효율적인 마케팅 전략 수립

05 | KT

올레 추천
시스템 구축

프로젝트 기간

사업기간 : 약 1개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 고객 성향에 적합한 컨텐츠 추천 서비스를 제공하기 위해, 컨텐츠 구매 및 이용로그를 활용한 분석모델 구축

분석내용

고객 성향 및 관심사
지수 개발

구매/이용 로그 기반의 서비스 타입 별 이용 고객의 성향을 정의하고
App/Movie/Music을 이용한 고객성향에 따른 예측모델 개발

Double-Chained
Modeling 적용

서비스 타입별 중분류 카테고리 정보 기반의 성향 패턴을 도출하기 위해
군집화 결과를 Target으로 Decision Trees Algorithms 재 적용하는 기법 사용

기대효과

올레마켓/라이프쟈키 각 서비스 이용 고객의 성향 분석에 기반한 맞춤형 추천 정보를 제공

국내 프로젝트 사례집

공공 분야 (Public Sector)

- | | |
|----------------|----------------|
| 01 NIA | 06 서울시청 |
| 02 한국전력(KEPCO) | 07 한국지역난방공사 |
| 03 식품의약품안전처 | 08 한국도로공사 교통센터 |
| 04 행정안전부 | 09 근로복지공단 |
| 05 ETRI | |

01 | NIA

공공데이터 활용 지수 측정 및 진단 모델 개발 사업

프로젝트 기간

사업기간 : 5개월

활용 Tool

Python

배경 및 목적

- 공공데이터 포털에 개방된 데이터들의 활용성 진단 및 개선 필요
- 개방 공공데이터의 활용성 증대 및 이용자 만족도 제고를 위한 진단 모델 및 도구 필요

공공데이터 활용 지수 측정 및 진단 시스템 구축

활용성 지표 생성 및 진단	<ul style="list-style-type: none"> · 메타데이터에서 추출한 데이터를 바탕으로 사전에 정의된 5가지 활용성 지표를 생성 · 사용자가 업로드 한 공공데이터의 활용성 지표 진단 결과 산출
AI 학습 분석 기반 활용성 예측 모델 생성	<ul style="list-style-type: none"> · 공공데이터 포탈에서 16개의 분류체계로 나누어진 데이터를 통합하여 학습 및 분류체계별 학습 진행, 생성된 통합 모델과 분류체계별 모델로 분류체계별 성능 비교 · 포탈에 업로드 된 데이터 형식인 파일 데이터, API 데이터를 분류하여 활용 지수 측정 모델 생성, 다양한 알고리즘 별 성능 비교를 통해 가장 우수한 성능을 보인 알고리즘 선택 <ul style="list-style-type: none"> - 파일 데이터 : Catboost 단일 알고리즘 모델 사용 - API 데이터 : Catboost + LGBM 을 조합한 양상을 모델 사용 · 모델 학습 시 각 분류체계 별 RMSE 값이 작은 챔피언 모델 선정 후 학습에 반영 · 챔피언 모델로 선정된 통합 모델 및 분류체계 별 모델 피클 파일로 저장
모델 반영 및 평가	<ul style="list-style-type: none"> · 1개월마다 포탈에 업로드 된 신규 데이터 포함하여 모델에 통합하고 이를 기반으로 재학습 진행 후 서버 반영, 사용자가 원하는 경우 모델을 개별적으로 재학습 및 반영 가능 · 기보유 데이터 업로드 시 활용성 진단 결과를 측정하여 산출 · 신규 데이터 업로드 시 활용성 예측 결과를 측정하여 산출

기대효과

- 공공데이터 개방 심의 시 개별 데이터의 민간 활용성을 객관적으로 측정할 수 있는 시스템 구축
- 민간의 활용성이 낮은 개방데이터에 대한 활용성 진단을 통해 해당 데이터에 대한 활용성 개선

02 | 한국전력공사

전력설비 자산성능평가 시스템 구축

프로젝트 기간

사업기간 : 15개월

활용 Tool

IBM SPSS Modeler
Client/Server
Python

배경 및 목적

- 전력설비 생애주기 모니터링 및 교체 투자효과 관리를 지원할 수 있는 시스템 및 기능 부족
- 데이터 기반의 객관적인 설비 별 교체 대상 판정시스템 도입 추진 필요

자산성능평가 시스템 구축

Risk 기반 자산성능평가	<ul style="list-style-type: none"> Risk = PoF (고장확률) X CoF (고장영향비용) 자산별로 산출한 Risk값을 기반으로 등급을 부여하고 등급에 따라 교체 대상 여부 판단 	
PoF (고장 확률)	통계잔여수명	운영/고장 데이터를 활용해 와이블 분석 기반으로 잔여수명 산출
	상태잔여수명	자산의 내/외부 상태점검 결과인 진단 데이터를 활용하여 각 상태평가 항목별 잔여수명 계산 및 상태등급 부여
	운전잔여수명	부하데이터 활용해 자산이 운영됨에 따라 감소된 잔여수명 산출
<ul style="list-style-type: none"> 통계/상태/운전 잔여수명을 기반으로 자산의 최종 잔여수명 및 PoF(고장확률) 계산 최대 10년까지 자산의 고장확률 및 예상 교체시기 제시 		
CoF (고장 영향 비용)	<ul style="list-style-type: none"> 재무/신뢰도/안전/환경 측면에서 각각 영향비용을 계산하고 이를 합산하여 최종 CoF 산출 	

기대효과

- 전력설비 노후교체에 소요되는 투자비 절감 및 설비 투자 효율성 증진
- KEPCO에 최적화된 리스크 기반의 자산성능평가 시스템구축



03 | 식품의약품안전처

수입식품 위해도 분석 시스템 개발(1차)

프로젝트 기간

사업기간 : 11개월

활용 Tool

R, IBM SPSS Modeler

배경 및 목적

- 위해도를 판별하기 위해 수입하는 식품에 무작위 샘플링 정밀검사를 실시하지만, 적발률이 높지 않음
 - 위해도 적발률을 높이고, 실시간 판별이 목적이며, 신고서 내용에 대해, 사전에 이상치를 판별하여 검출하는 작업 필요
- ※ 위해도 : 식품을 포함한 모든 품목에 대해, 가진 독성이 얼마나 노출되는지를 통해 사람에게 미치는 위해성을 의미

분석내용

수입식품 위해도 예측모델	<ul style="list-style-type: none"> · 식품/수산물/축산물에 따라 위해도를 판별하기 위한 입력변수가 서로 다르기 때문에, 품별로 구분하여 위해도에 대한 데이터 패턴을 분석하고 각 예측모델을 개발함. · 머신러닝 알고리즘을 통해 위해도를 예측하고 위해도 적발의 근거제시를 위해 GLM모델 활용 · 분기별 모델을 재 학습하여 입력변수별 가중치를 갱신하고, 모니터링을 통해 모델 성능 관리 · 통관 사전 및 사후 단계에서 위해도 예측점수를 제공할 수 있도록 개발함. 사전 : 제조업소(현지실사)에 대한 검사 / 사후 : 수입업체(지도점검), 유통제품(수거검사) <p>※ GLM : Generalized Linear Model(일반화 선형 모델)</p>
신고서 이상치 검출 알고리즘 개발	<ul style="list-style-type: none"> · 과거 동일 수입업체/제조업체/용도/품목 별로 유사 정도를 이상치 탐지기법으로 산출하고, 유사도 기준을 통해 이상치를 판단하고 신고서 내용의 사전 경고를 전달해줌. <p>※ LOF ; Local Outlier Factor = 데이터의 거리와 주변 밀도를 고려한 이상치 탐지기법</p>
위해정보 경향성 분석	<ul style="list-style-type: none"> · 해외 위해정보 문서에서 텍스트 분석을 통해 위해 관련 키워드의 경향성을 분석하기 위한 기능 · 기존 위해 키워드(40개)에 추가로 유사어와 불용어 등의 사전을 구축하여 전처리를 적용 · 각 키워드별 출현빈도 및 비율을 산출하여 추이를 분석할 수 있도록 시각화 적용 <p>※ 불용어 : 텍스트 분석에서 의미가 없는 데이터를 제거하는 작업이 필요, 해당 단어들을 의미함.</p>

기대효과

- 실시간으로 위해도를 판별하여 수입식품에 대한 안전성을 높이고, 신고서 검토 등의 프로세스를 개선하여 사전 위험 감지 등 효율적인 업무 프로세스 구축

03 | 식품의약품안전처

수입식품 위해도 분석 시스템 개발(2차)

프로젝트 기간

사업기간 : 2개월

활용 Tool

R, IBM SPSS Modeler

배경 및 목적

- 수입식품 위해도 예측 모델의 설명변수(feature)가 민감하여 일부 수입화주에게 위해도 무작위 몰입 현상 발생
- 무작위 선정 근거에 대해 직관적으로 이해할 수 있는 시각화 자료 부재로 민원 대처에 대한 어려움 발생
- 특정 품목, 국가 등의 수입 건에 대한 동향을 한눈에 파악할 수 있는 기능이 없어 즉각적인 대처에 대한 어려움 발생

수입식품 위해도 예측 모델 고도화

- 특정 수입화주에게 몰입되는 현상을 제거하기 위해 동일 수입화주, 제조국가, 품목을 기준으로 과거 무작위 선별 이력 여부를 확인하여 위해도 점수를 조정하는 별도 프로세스를 개발
- 기존의 무작위 선정 제한 비율을 위해도 평균 점수를 바탕으로 5개 등급으로 나눠 위험도에 따라 차등된 제한 비율을 적용하는 것으로 고도화하여 타당성을 제고

위해도 무작위 점수 산정 근거 시각화

- 수입신고서의 위해도 점수가 산정된 근거와 위해도 무작위가 선별된 이유를 시각화 화면을 통해 제공함으로써 담당자의 업무 효율성을 제고
- 최근 3년간 수입된 수입신고서의 위해도 점수 분포 중 해당 수입신고서의 위치를 Rating Chart를 통해 제공
- 위해도 예측 모델의 설명변수(feature)별 표준화된 가중치 값을 Spider Chart를 통해 제공하여 타 수입신고서 대비 어떤 항목에서 높은 점수를 받았는지 직관적으로 제공

수입식품 동향분석

- 수입된 수입신고서의 데이터를 회귀분석 및 t-test를 통해 증감률을 확인할 수 있는 다양한 지표를 산정
- 품목, 국가 등의 다양한 기준에 따른 동향을 확인할 수 있는 별도 화면을 제공함

04 | 행정안전부

공공분야 빅데이터 분석

프로젝트 기간

사업기간 : 6개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

고용노동부는 근로자의 근로조건 보호 및 기준 확보를 위해 근로 감독 및 산업안전 감독을 실시하고 있으나 대부분 감독관의 경험과 직감으로 사업장을 선택하고 있으며, 감독 사업장 수에 비해 감독관이 턱없이 부족함

물동량 수요 예측 모델 개발

- 질병자 발생 현황, 감독 결과, 유해 위험성 평가 등을 분석하여 산업재해 취약지수 산정 기준 도출
- 취약 지수 기반 사업장 선정모델과 데이터마이닝 기법(연관성분석, 군집분석, 의사결정나무, 신경망 등)을 적용하고 결과 비교를 통한 우수모델 적용

기대효과

노동관계법 위반 및 산업 안전사고가 예상되는 사업장을 미리 예측하는 등 과학화 된 감독 사업장 선정 체계를 마련하여 위반 대상 사업장 선정 업무의 효율을 높이고 범국민적 근로환경을 개선함



04 | 행정안전부

온라인 물가 지수 개발 및 활용과제 발굴

프로젝트 기간

사업기간 : 3개월

활용 Tool

R, SAS Visual analytics

배경 및 목적

행정부적 빅데이터 공유, 분석 체계 마련 및 활용 표준모델 필요성이 대두됨에 따라 각 부처별의 빅데이터 활용 수요가 지속적으로 발생하여 안전행정부에서 프로젝트를 수행

온라인 물가지수 개발

- 공통기반 인프라를 활용하여 국내 온라인 쇼핑몰에서 판매되는 상품 중 통계청 CPI 품목에 해당하는 상품의 가격을 웹 크롤러로 크롤링하여 URL 데이터를 수집
- R을 이용하여 매일 온라인 물가지수를 생성하고 이를 바탕으로 분류별 지수, 사이트 지수, 대형마트 지수 등을 계산
- 데이터를 통계청과 상호 연계하여 시각화 화면에 나타냄으로써 통계청 담당자들이 매일 물가지수를 확인하고 기존 CPI와 비교할 수 있도록 함

빅데이터 활용과제 발굴

- 통계청 담당자와의 인터뷰를 통해 외부 지표와 외부 데이터를 활용하는 방법론을 토대로 ‘빅데이터를 활용한 광역 및 위성도시의 일일 이동인구 분석’ 과제 도출
- 빅데이터를 활용한 광역 및 위성도시의 일일 이동인구 분석을 통해 이동 인구를 빠르게 파악함으로써, 교통정책, 축제 예산 산정 등의 정책에 활용하고자 함



05 | ETRI

우체국 물류 예측 모델 개발

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

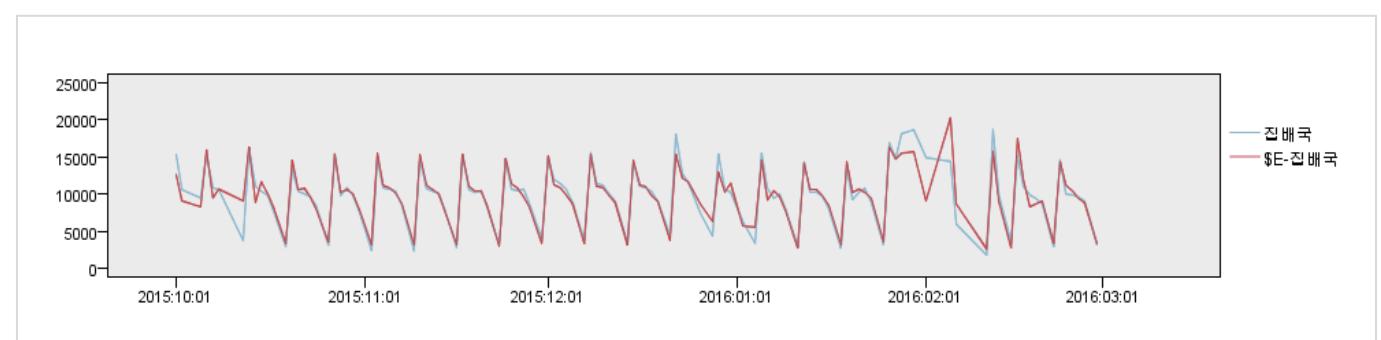
우체국 물동량 관련하여 예측이 필요한 시점에서 주요 급지(우체국 택배가 많은 지역)의 물동량과 택배 차량의 효율적인 배분을 통해 물동량을 관리하고자 함

물동량 수요 예측 모델 개발

- 주요 급지별 평균/일주일/한 달/명절과 같은 캘린더 변수들을 통해 시계열적인 변화량 감지
- 산지 및 특정 지형에 관해서는 예측이 어려워 중점을 두지 않음
- 시계열 모델과 회귀분석을 통하여 예측 모델 개발

기대효과

- 캘린더 변수를 활용하여 물동량이 급감하는 구간과 급증하는 구간의 명확한 차이를 통해 탄력적인 물동량 관리가 됨
- 개발 당시 사람을 통해 예측하던 것을 정확한 데이터를 통해 예측을 시작



06 | 서울시청

빅데이터 공유·활용 플랫폼 고도화

프로젝트 기간

사업기간 : 6개월

활용 Tool

IBM SPSS Modeler
QGIS

장애인 이동권 분석 – 시각장애인 콜택시

- 콜택시의 수요 및 공급 패턴 분석을 통해 현황을 파악
- first/last call의 밀집 지역을 권역으로 분류하여 권역 별 first/last call과 각 주차장까지의 최단거리 분석을 통해 권역별 적정 차고지 방안 수립
- 가용 자원을 기준으로 시간대별 수요 대비 공급이 충족될 수 있는 적정 차량 공급량을 분석

장애인 이동권 분석 – 장애인/노약자 무료 셔틀버스

- 서울시 및 13개 자치구에서 운영하는 장애인/노약자 무료 셔틀버스에 대해 이용 현황을 파악
- 장애인/노인의 거주/유동인구 데이터를 기반으로 인구 분포는 많으나, 기존 노선이 커버하지 못하는 사각지역을 분석
- 이용자수 예측 모델을 통해 사각지역에 정류장 신설 시, 이용자수를 예측하여 대안노선 방안 수립

결핵환자 발생 및 유형 구분에 따른 지역의 특성요인 분석

질병관리본부 결핵정보 통합관리 시스템(TBnet) 결핵 신고환자, 국민건강보험공단 진료 데이터를 통해 서울시 동단위 지역의 인구특성별(성/연령/소득수준) 결핵 발생 현황과 치료 현황을 파악하고, QGIS를 활용한 지역별 환자 분포를 통해 집중 관리지역을 선정



07 | 한국지역난방공사

열수요 예측 모델 개발

프로젝트 기간

사업기간 : 12개월

활용 Tool

IBM SPSS Statistics &
IBM SPSS Modeler

배경 및 목적

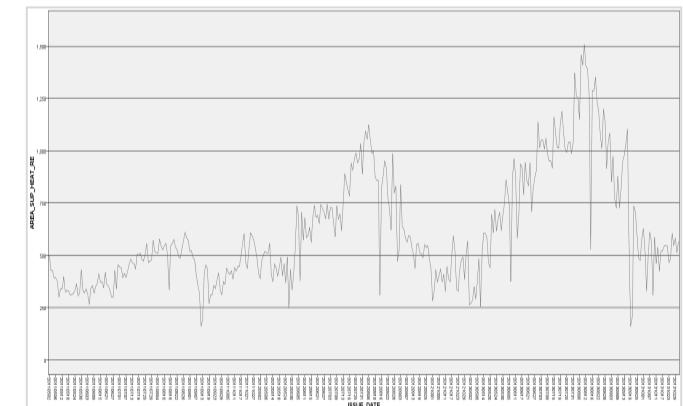
- 지사 단위의 주관적인 공급량 판단 및 설비 이용 등의 비효율적인 운영과 향후 연계 배관망이 추가 구성됨에 따른 전체 네트워크 통합 운영의 필요성 대두
- 열 공급 부하 예측 정확도를 개선하여 에너지 이용 효율 향상 및 수익 극대화 전략 도출

열수요 예측 모델 개발

- 주요 인자를 도출하기 위해 용도별 열수요 특성을 분석
- 시계열 모델을 추가 적용하여 지사별/시간대별 모델을 적합하였으며, 시계열/회귀 모델에서 설명하지 못하는 현상을 보정하기 위해 캘린더/온도/수요실적 효과에 따른 보정 모델을 개발함

기대효과

- 지사별 열수요 특성 분석에 기초한 열 공급 예측 프로그램 보완 기대
- 열공급부하 예측 개선에 따른 비용 절감
(일평균 오차율 1.34%p (15%) 개선 시 전체 동절기 약 359백만원 절감)



08 | 한국도로공사 교통센터

교통예보 지원 시스템 기능개선

프로젝트 기간

사업기간 : 6개월

활용 Tool

IBM SPSS Statistics 20.0 Client/Server

배경 및 목적

예측정보의 다양성 확보, 예측 정확성 제고 및 예보 지원 시스템 개선

예측 방법론 개선

소요시간 예측 시 기준에는 동일 요일의 최근 20주 데이터를 활용하였으나, 당일 예측 교통량을 기준으로 ± 10 만 대(or 20만 대) 교통량에 해당되는 과거 10주 데이터를 활용함으로써 정확성 증대

민자노선 및 버스 기준 소요시간 예측

서울/광주, 부산 구간 소요시간 예측 시 민자노선인 천안/논산선, 대구/부산선 경유 경로에 대한 소요시간 예측 모델 개발 및 고속버스 기준의 5대 도시 소요시간 예측 모델 개발

버스 전용차로 소요시간 비교정보 제공

- 버스 전용차로(서울↔오산, 신탄진)와 일반차로의 소요시간 비교 정보 표출을 통한 대중교통 이용 유도 목적
 - BUS VDS 정보를 이용하여 버스 전용차로의 평균 속도 및 코존 단위의 정체 길이 예측 모델 개발

실시간 교통상황 반영 소요시간 예측

과거의 정체 상황과 현재 정체 상황을 비교하여 정체가 풀리거나 증가한 만큼을 예측 소요 시간에 가감하여 예상 소요시간을 조정 및 실시간 교통상황 반영



09 | 근로복지공단 FDS 분석 시스템 구축

프로젝트 기간

사업기간 : 6개월

활용 Tool

IBM SPSS Modeler 14.2

Client/Server

배경 및 목적

산재보험 재해발생 대상자들에게 부정혐의 점수를 부여하여 사기 적발 가능성이 높은 대상자 추출 및 제공

사기 적발 스코어링 모델

산재보험 부정수급 대상자를 추출하여 부정혐의 점수를 제공하는 스코어카드 모델 개발

Risk Factor Pool 구축

자료수집, 후보Factor 검토, Factor정의, 최종Factor 결정의 4단계를 거쳐(잠재 Factor 까지 포함한) Risk Factor Pool 구축

이상징후 모니터링

이상징후 지표를 선정한 후, 특정 기간별 추이를 모니터링하고 정해 놓은 임계치를 벗어나는 경우 Alert를 제공

연관성분석(Link Analysis)

- 피해자 중심의 전화 통화 내역을 통해서 용의자를 추출
 - 방향성 데이터가 존재하지 않기 때문에 연결형태와 강도를 가지고 탐색

총 의료기관별 전문 분야 [504093]				최종등록일
시행일	전공	의료기관명	목록	선택하기
해당항목제거	국립현암병원	의료기관명		
0 조기 대체				
기사	의료기관번호	의료기관명	의료기관명	전문분야
서울기전분부	100102	글로벌한생과의원	지정의원	2
서울기전분부	100103	보건기전분부	보건영업	4
서울기전분부	100105	한국기전분부의원	의원	18
서울기전분부	100106	서원병원	급상생동	67
서울기전분부	100109	한국기전분부의원	의원	35
서울기전분부	100115	서원불교성모병원	의원	2
서울기전분부	100125	한국기전분부의원	의원	4
서울기전분부	100130	한국기전분부의원	의원	17
서울기전분부	100160	한국기전분부의원	의원	9
서울기전분부	100184	한국기전분부	의원	2
서울기전분부	100187	한국기전분부의원	의원	13
서울기전분부	100195	한국기전분부의원	의원	2
서울기전분부	100203	한국기전분부의원 풍물동	의원	28
서울기전분부	100210	한국기전분부의원	의원	3
서울기전분부	100245	한국기전분부	급상생동	67
서울기전분부	100246	한국기전분부의원	의원	3
서울기전분부	100310	한국기전분부의원	의원	15
서울기전분부	100448	의료법인한국기전분부의원부의과의원부임상	교대영업	25
서울기전분부	100535	한국기전분부의원	의원	4
서울기전분부	100603	한국기전분부의원	의원	5



국내 프로젝트 사례집

HR (Human Resources)

01 현대자동차

01 | 현대자동차

우수 승진자를 예측하기 위한 승진적합도 분석모델 개발

프로젝트 기간

사업기간 : 2개월

활용 Tool

IBM SPSS Modeler

배경 및 목적

- 연구소 내의 각 팀에서 추천하는 임직원을 승진시켰을 때, 데이터기반의 분석모델을 통해 해당 직원이 실제 우수한 직원인지 판별하여, 무분별한 승진추천을 관리하고 투명성과 신뢰성 제고 목적
- ※ 승진대상 임직원 : 연차진급 및 책임진급(두가지만 대상)

분석내용

타겟 정의	<ul style="list-style-type: none"> 우수승진자를 정의하기 위한 다양한 관점의 타겟 정의 <ul style="list-style-type: none"> - 우수고과자 : 승진 후 당해년도 상/하반기 고과평가 중 최고등급 1회 이상 부여 받은 대상 - 승진만족자 : 승진 후 현재 팀장의 만족도 설문조사 결과 1순위를 부여 받은 대상
분석모델	<ul style="list-style-type: none"> 데이터 수집 : <p>승진적합도 분석모델을 개발하기 위해, 과거시점(17~19년도) 승진대상자 및 승진자를 대상으로 인적정보(성별, 연령, 거주지, 맞벌이여부, …), 조직정보(조직장연령, 연령차이, 조직 경험 횟수, 직무 경험 횟수, …), 자격증/논문/특허 정보, 웹로그 접속 정보, … 등 다양한 데이터를 수집 및 활용</p> 변수선택 : <p>IV(Information Value)를 통해 정의한 타겟변수별 입력변수의 관계를 통해 IV기준 0.5/0.1/0.2 3가지로 변수를 선택함</p> 모델개발 : <p>랜덤 포레스트 알고리즘을 사용하여 수집된 데이터별, 선택한 변수기준을 가지고 최적화된 파라미터를 적용하여 학습 및 검증을 통해 성능 평가하여 최종 모델 개발함</p>

기대효과

- 팀 내의 추천에 의해 주로 이루어지던 승진세션을 객관적인 데이터로 우수승진자 예측으로 무분별한 승진세션 관리 및 신뢰성과 투명성 제공
- 승진을 포함한 채용 등 HR영역의 분석모델 적용을 통한 고도화 가능

국내 프로젝트 사례집

Thank you

서울특별시 강남구 언주로 620 현대인텔렉스빌딩 10층(5, 11층)

TEL 02.3467.7200 FAX 02.852.3777

www.datasolution.kr