

머신러닝과 딥러닝을 이용한 고객만족 예측모델 개발

Development of customer satisfaction prediction model using deep learning

김 계 수*

Kim Gye-soo

ICT기반 4차 산업혁명과 코로나 19 확산으로 제조업과 서비스 영역이 점차 모호해지는 서비타이제이션(sevritization)과 빅블러(big blur)현상이 가속화되고 있다. 이러한 상황에서 데이터 기반한 고객요구와 기대를 명확하게 분석하고 이를 의사결정에 반영할 필요성이 제기되고 있다. 산업계에서 실제 머신러닝과 딥러닝 기반 인공지능을 실무에 적용하는 사례가 등장하고 있으나 학계에서는 아직 이에 대한 연구가 미흡하다. 본 연구는 A커피전문점의 고객관련조사 데이터를 머신러닝과 딥러닝을 이용하여 고객만족 예측모델을 개발하는데 있다. 또한 고객만족 예측모델 개발 프로세스를 제시하는데 있다. 종속변수가 이변량(0, 1)인 경우의 예측에 사용되는 로지스틱회귀분석기법을 적용하고 이어 본 연구에서는 고객향후 방문 여부(0, 1)를 종속변수로 하고 학습 데이터와 검증데이터의 비교를 통해 모델예측 성능을 비교하기로 한다. 다음으로 구글 브레인팀이 개발한 딥러닝(deep learning)과 머신러닝(machine learning) 전문 라이브러리인 텐서플로(TensorFlow)를 이용하여 인공지능 분석을 실시하기로 한다. 분석결과 고객만족 예측모델의 정확도(Test accuracy)는 71.8%이었다.

핵심용어 : 서비타이제이션, 빅블러, 인공지능, 머신러닝, 딥러닝, 고객만족 예측모델, 텐서플로, 정확도.

* 교신저자, 세명대학교 경영학과 교수, E-mail: gskim@semyung.ac.kr

본 연구는 이 논문은 2021학년도 세명대학교 교내학술연구비 지원에 의한 연구임.

투고일 2022.03.03

수정일 2022.04.07

게재일 2022.04.30

1. 서 론

제조업과 서비스업의 미래는 데이터 기반 디지털 혁신에 길이 있다. 디지털 혁신을 통한 시장 참여자가 계속 등장하면서 덩치 큰 조직이 힘을 발휘하던 시대가 아니라 데이터 기반으로 고객에게 지속적인 가치를 제공하는 기업이 생존하는 시대가 도래한 것이다.

디지털 전환의 혁신속도가 곧 미래경쟁력이다. 현재 기업들이 직면하고 있는 현안은 기후위기 대응과 디지털 혁신이라고 할 수 있다. 탄소감축이 불가피한 자동차, 석유화학 등 업종에서 선제적인 대처가 필요하다. 빅데이터, AI, 메타버스 등의 기술을 통한 스마트 제조 또는 지능화 제조가 이루어지도록 해야 한다.

제조업과 서비스의 융합이 가속화되는 서비타이제이션(servitization) 시대와 제조업과 서비스업의 영역이 모호해지는 빅블러(big blur)상황에서 제조부문과 서비스부문 경쟁력 강화 필요성이 제기되고 있다. 서비스 기업들은 머신러닝과 딥러닝의 AI 핵심기술을 적극 이용하고 있다. 상품과 서비스의 생산과 유통 소비의 가치사슬단계를 연결하고 지능화하려고 노력하고 있다. 정보통신기술(ICT)발달로 빅테크, 유통, 통신업 등과 결합이 가속화되고 있다. 고유업종의 경계가 허물어지는 상황에서 고객 라이프 스타일의 데이터화를 통한 기회영역 발굴이 경쟁력인 시대가 도래하였다.

2020년에 발생한 코로나(COVID19)는 언택트(untact) 또는 온택트(ontact) 서비스를 가속화시키고 있다. 기업과 조직을 둘러싼 환경중에서 COVID-19는 조직구조 변화에 영향을 미치고 있다. COVID-19는 공장폐쇄에 따른 공급사슬 단절을 가져와 생산중단과 무직휴직 및 해고를 양산하고 있는가 하면 경제활동 저하, 자산가치 감소,

그리고 소비심리지수 하락을 유발하고 있다. COVID19에 따른 경기침체도 문제이지만 더 중요한 것은 위기의 고착화가 더욱 문제이다. COVID19는 소비자의 변화, 업무환경과 일하는 방식, 산업 및 경쟁구조 등에서 다양한 영향을 미치고 있다. COVID19는 국내 온라인 유통시장에도 영향을 미쳐 최소 전세계 온라인화율은 점차 증가하고 있는 추세이다.

서비스 전달과정에서 최종 1인 소비자에게 개인화된 서비스를 제공하는 것이 엔드투엔드(End to End) 서비스 완결이자 경쟁력 원천이라고 할 수 있다. 기업 경쟁력은 고객 거래 데이터 분석에 기반한 고객요구 파악과 서비스 전달력에 있다. 고객 데이터 축적으로 알고리즘 등 기술을 이용한 서비스 개발은 무엇보다 중요하다. 신세계 그룹의 SSG.COM 경우는 데이터 기반 의사결정으로 개인화 서비스 제공을 위해서 고객의 거래실적 분석은 물론 컨텍스트 인지 사용자 거래(context-aware ux)를 예측형 고객 접근모델을 개발하여 실무에 적용하고 있다(노유정, 박한신, 2021). SSG.COM은 고객을 더 잘 이해하고 더 잘 팔고 더 잘 배송하고 더 효율적으로 일하는 것이 중요해진 상황에서 인프라와 사람을 확충하고 'DT(Digital Transformation) 본부'를 신설해 인공지능과 IT 인력을 대거 영입하였다(노유정, 박한신, 2021).

신한카드의 경우는 자사의 소비 데이터로 글로벌 결제 네트워크 비자와 함께 만든 '신한 그린인덱스'를 개발해 ESG 컨설팅 사업을 실시하고 있다. 신한 그린인덱스는 소비자가 산 물건의 탄소 배출 정도를 지수화한 것이다(이인혁, 2022). 예를 들어, 소비자가 편의점에서 지출한 1,000원당 얼마만큼 이산화탄소가 배출됐는지, 운송 수단을 이용했다면 km당 탄소 배출량은 얼마인지 등을 산출한다. 신한카드는 데이터 기반 ESG 컨설팅 사

업 진출을 위해 적용 데이터 축적, 업종 선별, 알고리즘 설계 등 전반 전략을 수립한다. 카드 제휴사와 공동으로 친환경 가맹점을 고객에게 추천하는 등의 마케팅을 펼치거나 리워드를 주고 ‘착한 소비’를 유도하는 친환경 생태계 구축 방안을 모색하고 있다. 신한카드 회원에게 개인별 맞춤형 탄소배출 현황을 알려주는 ‘그린 리포트’ 서비스도 제공하고 있다.

디지털 전환(DT, Digital Transformation)의 핵심은 인공지능 기술을 활용하여 고객경험을 새롭게 디자인하는 것이다. 현재까지 등장한 인공지능 자연어 처리 모델 중 가장 강력한 성능을 보이며 다양한 분야에 혁신을 가져올 것으로 예상되는 것이 GPT-3이다(<https://openai.com/blog/openai-api/>). GPT-3는 1,750억 개의 매개변수, 입력된 단어 4,990억 건, 350억Gbyte GPU 메모리 등을 사용하며 언어모델에 소요된 비용 1,200만 달러를 상회한다. GPT-3는 기존 AI와 다르게 많은 분야에 답을 할 수 있다. 단일 문장을 넘어 언어 문맥 파악이 가능하고 창의적인 대화를 내놓은 수준이 되었다. 전문 기자가 작성한 뉴스 기사와 구분이 어려운 수준의 뉴스 기사를 작성할 수 있다. 실제로 GPT-3와 대화를 나눠보면 GPT-3는 인류의 어리석음과 사랑에 대해 얘기하고 거짓말도 하는 수준이다.

본 연구에서는 머신러닝과 딥러닝을 이용한 고객만족 예측모델 개발에 대한 기본 전략을 제시하는데 있다. 세부 연구 목표는 다음과 같다.

첫째, 머신러닝과 딥러닝 관련 이론연구를 통해서 머신러닝과 딥러닝 관련 지식을 체계화하는데 있다. 고객과의 거래실적인 데이터를 본격적으로 분석하기에 앞서 머신러닝과 딥러닝의 개념과 학습 알고리즘을 이해하는 것이 중요하기 때문이다.

둘째, 한 기업의 고객관련 데이터를 확보하여 머신러닝과 딥러닝 알고리즘으로 고객만족 예측

모델을 제시하고 모델의 타당성을 확인하는 방법을 소개할 것이다. 본 연구에서 고객만족의 결과 변수인 재이용 의향 여부 문항을 이용하여 로지스틱 회귀분석결과에 기반한 결론과 시사점을 제공할 것이다. 최종 소비자에게 개인화된 서비스(n=1)를 제공하는 것이 중요한 시점에서 데이터 관찰, 데이터 분석, 개인화된 서비스 제공은 초우량 서비스 기업의 핵심역량이다.

셋째, 본 연구에서 제시하는 알고리즘과 결과분석이 현업과 실무분야에 도움이 되기를 간절히 바라며 이를 통해서 기업의 경쟁력 강화에도 도움이 되기를 희망한다. 또한 딥러닝과 머신러닝 연구에 관심이 있는 연구자에게 유익한 내용이기를 바란다. 따라서 본 연구에서는 연구자가 사용한 데이터와 알고리즘을 부록에 제공함으로써 학습자나 실무자가 실습할 수 있도록 할 것이다.

2. 이론적 배경

2.1 머신러닝

통계학(Statistics)과 머신러닝(ML: Machine Learning)은 유사한 듯 하지만 목적에서 차이를 갖는다. 통계학은 기존 데이터에 대한 설명에 주안점을 두고 체계적인 의사결정을 위한 학문이라면 머신러닝은 인공지능의 토대가 되는 것으로 다양한 데이터 분석을 통해서 ‘예측’에 무게중심을 둔다(김계수, 2020). 통계학은 검증에 무게중심을 두고 발생한 현상을 설명하는데 무게 중심을 둔다. 반면에 머신러닝은 수집된 데이터를 분할하여 학습용으로 모델을 개발하고 테스트용으로는 모델을 검증하여 모델의 예측력을 높이는데 사용된다. 서포트 머신러닝의 실용적인 가치는 거대하고 다양하다. 제조업에서는 머신러닝을 통해서 유

지보수 시점을 찾아낼 수 있고 고객에게 새로운 가치를 창출하는 방법을 탐색할 수도 있다. 서비스업에서는 방대한 데이터를 토대로 생각지도 못한 환경변화, 고객의 요구 변화 등의 해안을 찾아낼 수 있다.

머신러닝은 사람이 처리 불가능한 단시간에 대량의 데이터에서 자동으로 정확한 결과를 얻을 수 있다는 점에서 장점을 갖는다(이시카와 아키히코, 2020). 데이터 과학자들은 머신러닝 기법으로 서포트 벡터 머신(SVM: Support Vector Machine)을 이용한다. 서포트 벡터 머신은 복잡하고 강력한 머신러닝으로 알려져 있다. 서포트 벡터 머신은 1990년에 AT&T 실험실에서 블라드미르 베프닉(Vladimir Vapnik)과 동료에 의해서 개발되어졌다(테런스 J. 세즈노스키, 2018).

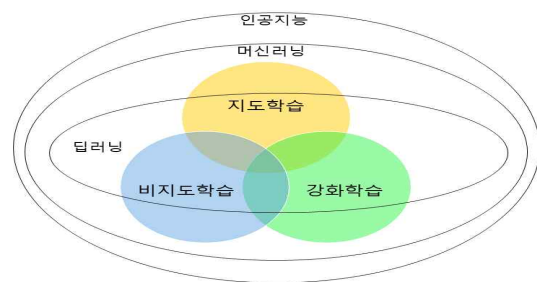
머신러닝은 얻은 데이터로부터 학습을 통해 지능이 요구되는 문제를 해결할 수 있는 컴퓨터 프로그램과 관련된 인공지능 영역을 말한다. 머신러닝에는 지도학습(supervised learning, SL), 비지도학습(unsupervised learning, UL) 그리고 강화학습(reinforcement learning, RL)이 있다.

지도학습은 정답 레이블이 붙은 데이터를 이용하여 머신러닝이 작동하여 학습하고 모델을 만들어 내는 것이다. 지도학습에 해당하는 대표적인 예는 회귀(regression)와 분류(classification)이다. 반면에 비지도학습은 정답 레이블이 붙지 않은 데이터를 이용하여 모델을 개발하는 것이다. 비지도학습의 대표적인 것은 주성분분석(PCA: Principal Components Analysis)과 군집화 분석(cluster analysis)이 해당한다.

강화학습은 시행착오를 통해서 학습한다. 학습에서는 사람이 데이터에 대한 라벨링을 할 필요가 없으며, 데이터를 수집하거나 이와 관련된 작업을 할 필요도 없다. 강화학습의 목표는 그저 행동하기이기 때문이다.

데이터사이언티스트는 머신러닝을 통해서 분류, 회귀분석, 탐지, 데이터 이상치 발견, 모델 과적합 발견으로 좋은 예측모델 생성, 많은 변수와 10,000회 이상의 학습을 신속하게 진행, 자동적으로 데이터 이상치 탐지 등을 할 수 있다.

[그림 2-1] 머신러닝, 딥러닝, 인공지능



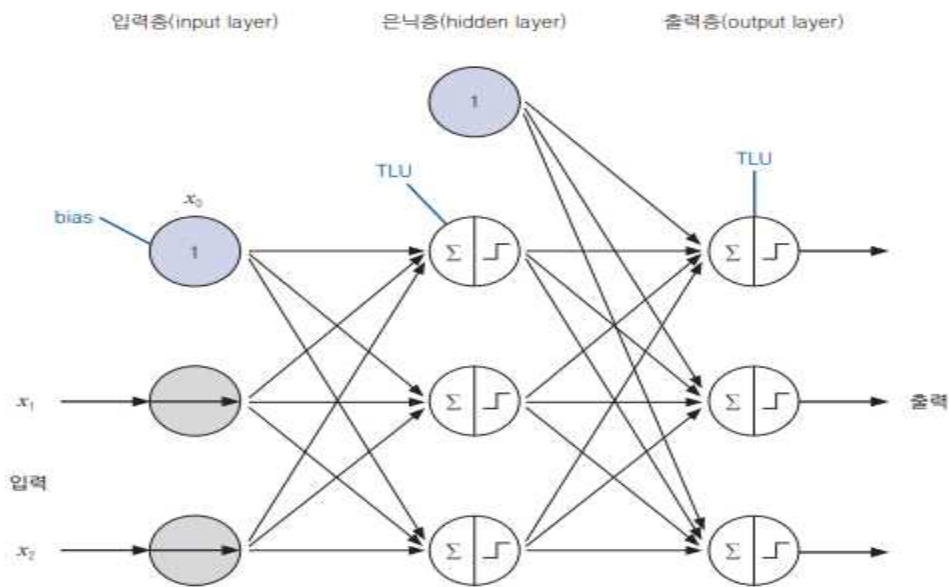
2.2 딥러닝

딥러닝은 수학과 컴퓨터공학, 신경과학에 뿌리를 두고 있는 머신러닝의 한 기법이다(테런스 J. 세즈노스키, 2018). 딥러닝(deep learning)은 특정 형식의 인공 신경망(ANN)을 사용한다. 딥러닝이라는 용어는 소위 심층 인공 신경망(ANN)에 대한 훈련을 의미한다. 우선 샘플 데이터를 통한 훈련작업이 필요하다. 훈련작업 이후에는 ANN을 해당 작업에 사용할 수 있다. 딥러닝에서 훈련된 ANN의 사용하는 것을 추론(inference)이라고 한다. 추론이 진행되는 동안 ANN은 학습된 규칙에 따라 제공된 데이터에 대한 평가 결과를 다시 보고한다. 평가 결과는 입력 이미지에 결함이 있는지 아니면 오류가 없는 객체를 나타내는데 대한 추정이 해당한다.

인공신경망(ANN)은 서로 연결된 뉴런으로 구성된다. 인공신경망에서 뉴런의 계층은 입력 계층, 은닉 계층, 출력 계층으로 이루어진다. 뉴런과 연결(link)은 매트릭스와 유사하다고 할 수 있다. 링

크 매트릭스는 입력 매트릭스 개별 값과 결과 매트릭스의 값 사이의 연결을 포함한다. 딥러닝을 이용한 인공지능에서 신경망인 파라미터가 사람의 뇌처럼 학습하고 판단한다. 따라서 파라미터 수가 인공지능망의 성능을 결정한다.

수이다. 입력층에서 어느 입력변수가 중요한지를 나타내는 가중치가 부여되고 은닉층에서 연결선을 통해 들어오는 값을 가중해 사용한다. 그리고 가중된 입력을 모두 합산하여 활성화 함수를 입력한다. 활성화 함수의 출력값은 노드의 출력이며



[그림 2-2] 인공신경망 구조

각 비즈니스 분야에 응용할 경우 전문가 수준 이상의 의사결정을 돕는 AI 인간을 개발하는 게 가능하다. 연결 매트릭스의 값에는 각 연결의 가중치가 포함한다. 입력값과 논리 매트릭스의 값에 가중치를 반영하면 결과 매트릭스의 개별 값이 생성된다. 이 인공신경망은 입력 및 출력 계층 뿐만 아니라 입력 및 출력을 위한 가시적인 계층 사이에 존재하는 수백 개의 추가적인 은닉 계층으로 구성한다. 숨겨진 계층의 결과 매트릭스는 다음 계층의 입력 매트릭스로 사용된다. 이 경우에는 마지막 계층의 출력 매트릭스에만 결과가 포함한다. 인공신경망은 입력과 출력을 연결한 합

이 값은 노드와 연결된 노드에 전달된다.

인공 신경망(ANN)은 생물학적 신경망의 수학적 모델이다. 인공 신경망이 훈련할 때 초기 초점은 무작위로 설정된다. 이후 샘플 데이터가 서서히 추가된다. 학습 규칙은 입력 데이터 및 예상 결과에 따라 관계의 가중치를 조정하는 데 사용된다. 결과에 대한 평가의 정확성을 의미하는 ANN의 궁극적인 효과는 훈련에서 사용되는 데이터에 큰 영향을 받는다. 훈련 내용에 변동성이 높은 학습 데이터가 많이 포함될수록 추론에서 더 정확한 결과를 얻을 수 있다. 매우 유사하거나 반복적인

데이터를 사용하여 훈련을 수행하는 경우, ANN은 예시 데이터와 다른 분야의 데이터를 추정할 수 없게 된다. 이 경우를 ANN에서 과적합(overfit)이라고 한다.

3. 고객만족 예측모델 개발과 분석

3.1 고객만족 예측 모델 개발

모델은 현실세계의 축소판이다. 모델은 물리적 표현, 자연어 문장, 컴퓨터 프로그램, 수학 방정식 등으로 나타낼 수 있다(김진형, 2020). 예측모델은 데이터 기반하여 미래 문제를 미리 알아볼 수 있게 하는 것이다. 본 연구에서는 서비스운영관리와 마케팅 분야에서 자주 언급되는 서비스 패키지(Bordoloi et al., 2019) 와 고객만족이론(Oliver, 1997, Ostrom, Iacobucci, 1995)을 토대로 고객만족 이후의 성과변수라고 할 수 있는 재이용 의향 변수를 이용하여 고객만족 예측모델을 개발하고자 한다. 또한 소비자들의 욕구를 만족시키기 위해 고객들에게 기업이 전달하기로 약속한 가치 또는 이익과 혜택들의 집합인 가치제안(value proposition)에 관심을 갖고 관련 연구를 시작하였다. 스타벅스의 경우, 고객은 스타벅스 매장에서 커피를 중심으로 모이고 지속적으로 성장하면서 직원 및 고객의 번영을 공유하는 것이 가치제안이며 존재이유이다(하워드 슬츠, 조앤 고든, 2020).

고객이 바라는 가치는 양적(가격, 속도)일 수도 있고, 질적(디자인, 고객의 경험)일 수도 있다. 기업이 고객에게 제공하기로 한 약속, 이익, 혜택이 가치 제안이며 고객 가치 제안이 고객에게 제공하고자 하는 고객만족과 고객 재이용의 결정요인이 될 수 있다.

이런 점에서 4차 산업혁명시대에서 변화 모습을 보이기 위해서 인공지능을 이용한 고객만족 예측모델 개발은 다음과 같은 순서로 진행할 수 있다.

고객만족 예측 모델링에서 테마선정은 고객관련 문제 정의를 내리는 것이다. 테마선정에는 고객만족 예측모델에 관한 핵심성과요인(KPI, Key Performance Index)를 설정해야 한다. 고객만족 예측모델 업무 프로세스를 정립하는 것이 중요하다. 고객접점 부서나 현장에서 발생하는 고객관련 문제를 정의하고 고객관련 문제관련 내용을 시각화(모형화 포함) 또는 수식화해서 나타낼 수 있다. 본 연구에서는 커피전문점인 A기업의 고객만족 예측모델을 개발하기 위한 것이다. 본 연구에서는 A커피 전문점의 지속적인 이용 여부의 결정요인의 예측모델을 개발하는 것이다.

테마선정 이후 데이터셋을 준비해야 한다. 데이터셋은 고객관련 데이터로 조직이 보유한 데이터와 고객 거래 데이터를 사용한다. 본 연구에서 사용하는 자료는 A커피전문점에서 조사한 자료인 일부인 2021년 커피점의 만족도 자료이다.



[그림 3-1] 고객만족 예측모델 개발 프로세스

머신러닝과 딥러닝을 이용한 고객만족 예측모델 개발

<표 3-1> 2021년 A커피 전문점 만족도 자료

변수 설명	변수유형	척도
개인사항		
X1(거래연도)	질적변수	1. 1년 미만 거래 2. 1~5 3. 5년 이상
X2(회원가입여부)	질적변수	0=가입, 1= 가입안함
X3(음료 구입성향)	질적변수	0=커피, 1=일반음료
X4(거주지역)	질적변수	0=경기, 서울지역 이외지역 1= 경기 서울
X5(소득 수준)	질적변수	0=낮음 1=높음
고객만족 결정 주요 변수		
x6(제품품질)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x7(정보품질)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x8(주차장지원)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x9(Wifi속도)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x10(영업시간단축)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x11(제품다양성)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x12(서비스요원이미지)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x13(실내혼잡도)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x14(확신성)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x15(신제품출시빈도)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x16(주문처리프로세스)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x17(적극적인 ESG활동)	양적변수	0=매우 좋지 않음 10=매우 탁월함
x18(착석대기시간)	양적변수	0=매우 좋지 않음 10=매우 탁월함
성과/관계 측정변수		
X19 - 만족도	양적변수	0=매우 좋지 않음 10=매우 탁월함
X20 - 추천의도	양적변수	0=매우 좋지 않음 10=매우 탁월함
X21 - 구매의도	양적변수	0=매우 좋지 않음 10=매우 탁월함
X22 - 구매수준	양적변수	0~100%
X23- 향후 애용 의지	질적변수	0=하지 않을 것임 1= 적극적으로 할 것임
x24-개선 및 불만사항	텍스트형	

데이터 전처리는 머신러닝 프로젝트에 사용하기 위해, 데이터를 모델이 이해할 수 있는 형태로 변환하거나 정보품질을 높이는 일련의 과정이다. 데이터 전처리 과정은 데이터분석이나 고객만족 예측모델 개발 프로세스 업무는 전체 업무(총 100%)중 평균 80% 이상을 차지하는 것이 일반적

인 견해이다. 데이터 전처리과정에는 무응답치 처리, 과대 산포, 이상치 해결 등이 포함된다.

피쳐 엔지니어링은 머신러닝 모델을 만드는데 있어 필수작업이다. 머신러닝 모델이 좋은 성능을 내려면 반드시 피쳐 엔지니어링이 필요하다. 데이터에 하나의 트리 모델을 학습 시킨 후, 해당 트

리 모델의 성능이 낮은 부분을 보완하는 다른 트리 모델을 학습 시키는 방식으로 수많은 트리 모델을 순차적으로 학습시키며 성능을 개선하는 모델이다.

모델 설계는 복잡한 현실 세계에 존재하는 데이터를 단순화하여 컴퓨터 세계의 데이터베이스로 옮기는 변환 과정이다. 모델 설계는 데이터베이스에 저장할 데이터를 위한 데이터 모델을 만드는 프로세스를 의미하며, 이러한 모델의 데이터 객체 간의 연결을 나타내는 개념적 표현이다. 또한 데이터 모델링은 데이터의 시각적인 표현을 도와준다.

모델 학습은 머신러닝의 지도학습에서는 취급하는 데이터를 훈련 데이터와 테스트 데이터로 나누어 사용한다. 훈련 데이터는 학습에 사용되는 데이터이며, 테스트 데이터는 학습된 모델의 정밀도(precision)를 평가할 때 사용하는 데이터이다. 훈련 데이터(training data)와 테스트 데이터(test data)로 나누는 이유는 머신러닝은 미지의 데이터를 예측하는 것을 목적으로 하기 때문이다. 본 연구에서는 훈련데이터를 70%, 테스트 데이터를 30%로 할당해서 모델을 분석하고 검증하기로 한다(김계수, 2020).

모델 모니터링은 모델의 타당성(model validity)을 확인하는 것이다. 앞 모델학습 과정에서 할당한 훈련 데이터와 테스트 데이터의 교차타당성을 판단하고 모델 타당성과 정확성을 판단하는 지수를 사용하여 고객만족 예측모델의 정확성 여부를 진단할 것이다.

마지막인 모델 사용 단계에서는 고객관련 예측 모델을 실무와 현업에서 적용하는 것이다. 본 연구에서는 이미 연구결과를 커피 전문점에 공유하였으며 본 연구결과를 각 체인점포에 전파하고 체인관리자와 서비스 파트너에게 학습자료로 제공하였다.

3.2 데이터 분석

3.2.1 로지스틱 회귀분석

앞에서 언급한 것처럼 본 연구에서는 A커피점을 이용하는 고객을 대상으로 한 데이터의 일부인 694명에 대한 데이터이다. 고객만족 센터는 향후 지속 이용 의향(x23, 0=이용 의향 없음, 1=이용 의향 있음)을 결정하는 독립변수를 알아보기 위해서 로지스틱회귀분석(Logistic Regression Analysis)을 실시하였다. 로지스틱 회귀모형(logistic regression model)은 종속변수가 이변량(0 또는 1)을 갖는 질적인 변수일 경우에 사용된다. 이 점에서 다중회귀분석과 근본적인 차이점이 있다. 실무에서 이변량의 데이터 경우는 많이 발견된다. 예를 들어, 실생활에서 의사결정(예와 아니오), 건강상태가 양호하거나 양호하지 않은 경우(생존과 죽음), 고객들이 회사의 제품을 구매하는 경우와 구매하지 않는 경우, 성공기업과 실패기업, 제품이 불량인지 정상품인지, 고객이 이탈할 것인지 잔류할 것인지 분류, 내원고객이 질병이 있는지 없는지, 이메일이 스팸인지 스팸이 아닌지 등 분류를 해야 할 경우가 많다. 대부분의 다변량 분석에서 독립변수들이 정규분포를 하며, 집단간 분산-공분산이 동일하다고 가정하나 로지스틱회귀분석에서는 정규분포의 엄격한 가정을 만족하지 않아도 된다(Hair et al., 2019).

본 연구에서는 전체 표본을 대상으로 하여 로지스틱회귀분석을 실시한 다음 전체 추정회귀모델의 성능을 파악하였다. 다음 단계에서는 모델의 타당성을 확인하기 위해서 학습표본을 70%, 검증표본을 30%로 랜덤하게 할당하여 분석하였다. 마지막 과정에서는 인공지능 모델을 적용하여 고객만족예측모델을 분석하였다.

3.2.2 분석결과

로지스틱 회귀분석 결과, 추정회귀식은 다음과 같다.

전체 데이터에 대한 로지스틱 회귀분석을 실시하였다(알고리즘 명령문: 부록1). 분석결과는 다음과 같다.

<표 3-2> 로지스틱회귀분석 결과

Dep. Variable:	x23	No. Observations:	694			
Model:	Logit	Df Residuals:	680			
Method:	MLE	Df Model:	13			
Date:	Sat, 26 Feb 2022	Pseudo R-squ.:	0.4258			
Time:	18:36:24	Log-Likelihood:	-272.69			
converged:	True	LL-Null:	-474.93			
Covariance Type:	nonrobust	LLR p-value:	2.533e-78			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-28.8884	3.074	-9.399	0.000	-34.913	-22.864
x6(제품품질)	1.1148	0.117	9.520	0.000	0.885	1.344
x7(정보품질)	0.1963	0.256	0.767	0.443	-0.305	0.698
x8(주차장지원)	-0.0770	0.122	-0.633	0.527	-0.316	0.162
x9(Wifi속도)	0.0780	0.198	0.393	0.694	-0.310	0.466
x10(영업시간단축)	-0.4583	0.130	-3.523	0.000	-0.713	-0.203
x11(제품다양성)	2.4578	0.676	3.637	0.000	1.133	3.782
x12(서비스요원이미지)	1.2250	0.205	5.987	0.000	0.824	1.626
x13(실내혼잡도)	-0.3333	0.090	-3.711	0.000	-0.509	-0.157
x14(확신성)	0.1616	0.232	0.695	0.487	-0.294	0.617
x15(신제품출시빈도)	-0.0466	0.078	-0.601	0.548	-0.199	0.105
x16(주문처리프로세스)	-0.0598	0.189	-0.317	0.751	-0.429	0.310
x17(적극적인 ESG활동)	2.9274	0.693	4.224	0.000	1.569	4.286
x18(착석대기시간)	-2.9886	1.286	-2.324	0.020	-5.509	-0.468

$$\hat{x}_{23} = -28.8884 + 1.1148x_6 + 0.1963x_7 - 0.0770x_8 + 0.0780x_9 - 0.4583x_{10} + 2.4578x_{11} +$$

$$1.2250x_{12} - 0.3333x_{13} + 0.1616x_{14} - 0.0466x_{15} - 0.0598x_{16} + 2.9274x_{17} - 2.9886x_{18}$$

분석결과 관찰된 결과의 우도가 높을 때 추정회귀모형이 적합하다고 할 수 있는데 여기서는 변수를 포함시키지 않은 상태에서 상수만을 포함한 경우의 LL값이 -474.93임을 보여 주고 있다. 모형의 설명력(Pseudo R-squ.)은 0.4258로 많은 독립변수가 투입된 것을 고려하면 설명력은 높다고 할 수 있다. x6(제품품질), x10(영업시간단축), x11(제품다양성), x12(서비스요원이미지), x13(실내혼잡도), x17(적극적인 ESG활동), x18(착석대기시간) 등은 $p < \alpha = 0.05$ 에서 유의한 변수임을 알 수 있다.

여기서 로지스틱회귀계수(coef)가 양수인 경우는 해당변수를 1단위 증가시키면 재이용 가능성이 양의 값(+)으로 증가하고 음수인 경우는 해당변수를 1단위 증가시켰을 경우 감소를 나타낸다고 할 수 있다. 전체적인 로지스틱회귀모델 성능을 파악하기 위해서 모델 예측 성능에 대한 적합도를 알아보기 위해서 정밀도(precision), 재현율(recall), F값(f1-score), support(지지횟수), 정확도(Accuracy)를 확인하였다(김계수, 2020). 우선 이들 지표를 알아보기 위해서는 혼동행렬(Confusion matrix)의 값을 살펴보았다. 분석결과는 다음과 같다.

<표 3-3> 혼동행렬표

LogisticRegression() Result of Classification :					
	precision	recall	f1-score	support	
0	0.83	0.81	0.82	393	
1	0.76	0.78	0.77	301	
accuracy			0.80	694	
macro avg	0.79	0.80	0.80	694	
weighted avg	0.80	0.80	0.80	694	
Confusion matrix:					
[[319 74]					
[66 235]]					
Accuracy:					
0.7982708933717579					
ROC_AUC:					
0.7962178658077824					

모델 예측 성능에 대한 적합도를 알아보기 위해서 정밀도(precision), 재현율(recall), F값(f1-score), support(지지횟수), 정확도(Accuracy)를 이해하면 된다. 정밀도는 모델이 예측결과와 실제결과에서 '0'이라고 한 것 중에서 정말로 '0'인 경우가 해당하는 사례 비율이다. 재현율은 실제 '0'에서 모델이 '0'이라고 판단한 사례 비율이다. 정밀도와 재현율이 모두 높은 것이 좋은 예측모델이라고 할 수 있다. 또한 정밀도와 재현율의 조화평균이 f1을 이용할 수도 있다. 우선 이들 지표를 알아보기 위해서는 혼동행렬(Confusion matrix)의 값을 살펴보면 된다. 참고로 혼동행렬을 분류표(Classification table)와 유사하다. 앞의 혼동행렬을 다음 표와 같이 정리할 수 있다.

<표 3-4> 혼동행렬

예측결과 실제결과		예측결과		퍼센트
		0 (재이용 없을 것임)	1 (재이용 예정)	
실제 결과	0 (재이용 없을 것임)	319 (TP)*	74 (FN)	81.2
	1(재이용 예정)	66 (FP)	235 (TN)	78.0
전반적인 퍼센트				79.8

혼동행렬에서 대각선 x_{ij} 는 제대로 맞춘 개수를 나타내고 그 이외의 성분은 다른 클래스로 오판하여 틀린 경우의 수를 말한다. 여기서는 오판 개수가 5개로 예측율은 대체로 있다고 볼 수 있다. 혼동행렬에서 판별결과에 대한 평가를 할 때 정밀도(적합률), 정확도, 재현율 그리고 F값을 보고 판단한다.

* 질병, 간호, 의학 연구분야에서는 TP(True Positive), FN(False Negative), FP(False Positive), TN(True Negative)를 나타냄.

· 정밀도(Precision, 0인 경우) =

$$\frac{TP}{TP+FP} = \frac{319}{319+66} = 82.8\%$$

· 정확도(Accuracy) =

$$\frac{TP+TN}{TP+FP+FN+TN} = \frac{319+235}{319+66+74+235} = 79.8\%$$

· 재현율(Recall, 0인 경우) =

$$\frac{TP}{TP+FN} = \frac{319}{319+74} = 81.2\%$$

· F값(f1 score, 0인 경우) =

$$\frac{2 \times Recall \times Precision}{Recall + Precision} =$$

$$\frac{2TP}{2TP+FN+FP} = 80\%$$

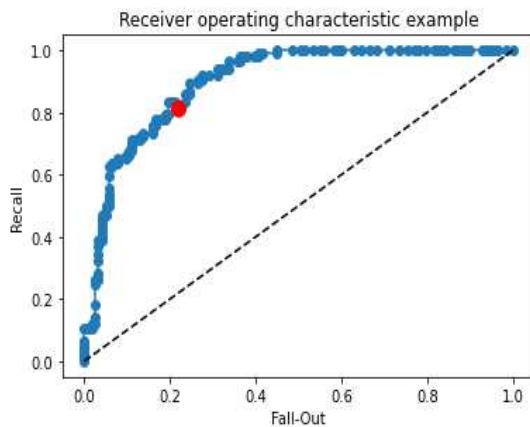
· 위양성율(fall-out) = $\frac{FP}{FP+TN} =$

$$\frac{66}{66+235} = 21.9\% (\text{낮을수록 우수함})$$

또한, 이진 분류(0, 1)시스템에 대한 성능평가에 사용하는 ROC_AUC(Receiver Operating Characteristic Area Under Curve)의 값을 살펴본 결과 79.6%였다. ROC_AUC는 분류에 모델이 얼마

나 적합한지를 나타내는 것이다. 제대로 예측이 안된 경우는 0%, 완벽하게 예측된 경우는 100%의 값을 갖는다. 추정로지스틱회귀모델에서 정밀도 82.8%, 정확도 79.8%, 재현율 81.2%, F1값 80%, ROC_AUC 79.6%로 나타나 대체로 분류가 잘 되었다고 평가할 수 있다. 또한 위양성율(fall-out)은 실제 양성 클래스에 속하지 않는 표본 중에 양성 클래스에 속한다고 출력한 표본의 비율을 나타내는데 21.9%로 낮은 편이라 성능이 우수하다고 할 수 있다.

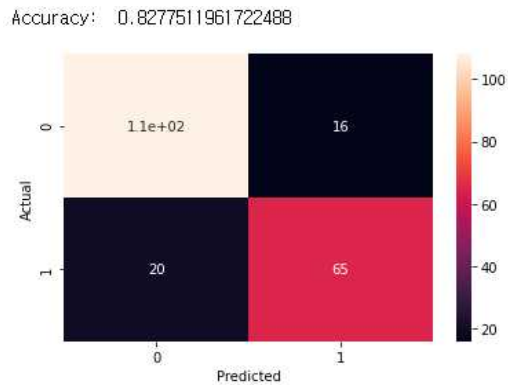
이어 ROC커브는 클래스 판별 기준값의 변화에 따른 위양성율(fall-out)과 재현율(recall)의 변화를 시각화한 ROC(Receiver Operator Characteristic) 커브를 그려보았다.



[그림 3-2] ROC Curve

ROC는 가장 이상적인 판정정도를 나타내는 것으로 가장 이상적인 것은 재현율(Recall)은 1이고, 양성율(fall-out)은 0인 경우이다. 로지스틱 회귀분석결과, 재현율(Recall, 0인 경우)이 0.812, 양성율(fall-out)이 0.219의 좌표가 붉은 원으로 찍혀 있다. 대체로 모델은 우수한 편이라고 할 수 있다.

또한 ROC 커브에서 모델의 평가가 좋다는 것은 커브의 밑면적 즉 AUC의 넓이가 넓을수록 우수함을 나타낸다.



[그림 3-3] 검정표본 혼동행렬

다음 단계에서는 모델의 타당성을 확인하기 위해서 학습표본을 70%, 검정표본을 30%로 랜덤하게 할당하여 로지스틱회귀분석을 실시하였다(알고리즘 명령문: 부록2). 검정표본의 혼동행렬표는 다음 그림과 같이 시각화할 수 있는데 여기서 예측 정확도는 82.8%임을 알 수 있다.

마지막 인공지능(딥러닝과 머신러닝) 모델을 적용하여 고객만족예측모델을 분석하였다(알고리즘 명령문: 부록3). 본 분석에서는 구글 브레인팀이 제작한 오픈소스 소프트웨어 라이브러리아자 응용프로그램 및 딥러닝(deep Learning) 전문 라이브러리인 텐서플로(TensorFlow)를 사용하였다. 텐서플로는 손쉽게 모델 제작이 가능하고 어디서든 강력한 머신러닝을 제작할 수 있고 연구를 위한 강력한 실험을 제공한다(<https://www.tensorflow.org/?hl=ko>). 본 연구의 인공지능 모델 분석에서는 인공지능 모델의 표현력을 높이기 위해 사용하는 활성화 함수로 시그모이드

(sigmoid) 함수와 ReLU 함수를 사용하였다. 시그모이드는 출력이 연속함수로 나타내는 경우이다. ReLU 함수는 입력값이 0보다 작으면 0으로 출력되고 입력값이 0보다 크면 그대로 출력하는 경우이다. 분석결과 검정표본 모델의 정확도(Test accuracy)는 0.718이었다. 본 연구에서 제시한 고객만족예측모델이 낮은 정확도를 갖는 모델이 아님을 알 수 있다.

4. 결론 및 시사점

4차 산업혁명의 시대에는 제조업과 서비스업의 고유영역이 점점 모호해 지고 있다. 이러한 빅블러(big blur)의 상황에서 서비스부문의 경쟁력 강화는 조직의 생존과 직결된다. 코로나 19 상황에서 온프라인과 오프라인을 통합은 가속되고 있다. 서비스 기업들은 머신러닝과 딥러닝의 AI 기술을 핵심동인으로 하는 상품과 서비스의 생산과 유통 소비 전과정에서 모든 것을 연결하려 노력한다. 이는 비즈니스 지능화를 통해서 고객관련 거래정보와 비즈니스 데이터화를 분석하여 고객만족 극대화, 기업생존을 영속화하려는 노력이라고 할 수 있다.

본 연구에서 서비스운영관리에서 중요시되는 서비스패키지와 마케팅분야에서 중시되는 고객만족결정요인을 토대로 A커피점 재이용에 관련 머신러닝과 딥러닝을 이용한 고객만족 예측모델 개발을 위해서 고객관련 자료를 확보하여 분석하였다. 분석결과 x6(제품품질), x10(영업시간단축), x11(제품다양성), x12(서비스파트너 이미지), x13(실내혼잡도), x17(적극적인 ESG활동), x18(착석대기시간) 등이 $p < \alpha = 0.05$ 에서 유의함을 알 수 있었다. A커피점에서 제품품질, 제품다양성, 서비스파트너의 이미지, 적극적인 ESG활동 등을 전개

하면 A커피점 재이용율은 증가할 것임을 알 수 있다. 또한 코로나19로 인한 거리제한에 따른 영업시간단축, 실내혼잡도, 착석대기시간 등은 불만요인으로 작용하여 이에 대한 관리가 이루어져야 함을 알 수 있었다.

또한 로지스틱회귀모델 예측 성능에 대한 적합도를 알아보기 위해서 정밀도(precision), 재현율(recall), F값(f1-score), support(지지횟수), 정확도(Accuracy)를 확인한 결과, 대체로 우수하였음을 확인하였고 인공지능(딥러닝과 머신러닝) 모델을 적용하여 고객만족예측모델을 분석한 결과 71.8%의 정확도를 보였다. 이러한 분석결과를 토대로 다음과 같은 시사점을 제시할 수 있다.

첫째, 인공지능을 이용한 고객만족 예측 모델을 개발한다고 하면 학계나 실무분야에서 어렵게 생각한다. 이유는 고객만족 예측 모델 개발은 외부 전문가로 구성된 조직을 만들고 믿고 맡기다 보니 프로젝트가 끝나면 조직 내부적으로 학습이 체화되지 못하기 때문이다. 또한 외부 전문가를 영입하지 않고 현업에 있는 기존 직원들에게 인공지능을 교육해서 요원을 양성하려는 경우는 인공지능 업무가 우선순위에서 밀려 경쟁력 축적이 안되는 경우이다. 이 경우에는 C-레벨(Chief level)에서 중역들이 직접 AI코딩도 배우고 실제 할 줄 알아야 한다. 이러한 경험이 급변하는 경영환경에 적합한 의사결정을 내릴 수 있다. 지속적인 내부교육을 활성화하고 임원을 포함한 모든 직원이 인공지능을 잘 알고 있을 때 인공지능을 응용한 고객관련 근본적인 변화와 혁신이 가능하다. 온·오프라인을 통합으로 차별적 소비자 경험을 제공하기 위해서는 고객에게 광적으로 집중하는 일체화된 팀 문화가 필요하다.

둘째, 고객경험이 중시되는 서비스나 유통에서 경험이나 직감이 아닌 데이터기반 체계적인 의사결정이 요구된다. 고객의 라이프 스타일이 바뀌는

지점을 데이터로 분석하고 관찰해 내면 필요 서비스를 발견해 낼 수 있다. 데이터는 성장의 원천인 원유이자 혈액이다. 데이터로 경험혁신을 이루고 혁신적인 비즈니스 모델을 개발할 수 있다. 오로지 데이터에 기반한 신속하고 정확한 실행 체계가 기업의 경쟁력을 끌어 올 수 있다. 고객경험 관리에서 데이터는 초개인화 서비스 제공을 위한 필요불가결한 요소이다. 고객의 소비 행태와 관련한 빅데이터에 근거해 주요 경영의사결정을 내리는 것이 필요하다. 모든 명확한 근거가 있어야 하는데 유일한 근거는 데이터이기 때문이다. 기업조직의 사업부 및 체인에서 축적한 데이터 즉 도메인 데이터 분석을 통한 디테일한 데이터 분석이 무엇보다 중요하다.

셋째, 탁월한 고객 경험을 제공하기 위한 서비스 전략에 치중해야 한다. 이를 위해서 고객경험 증시문화 및 고객중심의 비즈니스 프로세스 구축이 필요하다. 소비자들이 자유로이 온·오프라인을 넘나들며 시간을 보낼 수 있는 옴니채널 구축을 도입해야 한다. 고객 거래정보가 안전하게 지속적으로 관리되고 이를 기반으로 고객에게 인센티브를 제공하여야 한다. 특히, 고객관련 정보를 지속적으로 분석함으로써 비즈니스의 인사이트를 찾고 신제품개발 아이디어를 발굴하고 혁신과 개선의 아이디어를 탐색해야 한다. 데이터를 통해서 서비스 제공 전략상 중요한 전략상 가장 중요한 길목인 결정적 순간(MOT: Moment of Truth)을 관리해야 한다. 커피전문점의 경우는 제품품질과 서비스요원 친절이 가장 큰 무기인 것처럼 비대면 시대에는 인터넷, 스마트폰으로 인해 24시간이 결정적 순간으로 바뀌었다. 고객경험에 영향을 미치는 기업 홈페이지, 앱의 활용성부터 고객센터 직원의 전문성까지 모든 접점이 중요해지고 있다. 고객경험관리에서 관련 잠재 고객이 신규 고객이 되기까지 경험하는 경로가 SNS 광고인지 검색 광

고인지, 아니면 옥외 광고인지 정확히 분석할 필요가 있다.

본 연구에서 시도한 A커피전문점에 대한 로지스틱회귀분석과 인공지능망 분석 결과가 모든 서비스 영역에 적용하기는 무리가 있을 수 있다. 다만 향후에 본 연구방법이나 분석방법을 응용하여 타분야에서도 고객만족관련 경쟁력을 강화하는 계기가 되기를 기대한다. 또한 본 연구에서는 지면관계상 고객이 제시한 불만사항이나 제안사항에 관한 자료에 관한 텍스트마이닝 관련 결과를 제시하지 않았다. 이러한 내용은 후속 연구에서 제시할 것이다.

참고문헌

- 김계수(2020), Python을 이용한 데이터분석과 시각화, 한나래 아카데미.
- 김진형(2020), AI의 최강수업, 매일경제신문.
- 노유정, 박한신(2021), 강희석 대표 “경험이나 감만으로 유통업 할 수 없다”, 한국경제신문, A21면, 9월 15일.
- 이시카와 아키히코(지음), 박광수 옮김(2020), 파이션으로 배우는 딥러닝 교과서, 한빛미디어.
- 이인혁(2022), “신한카드, 빅데이터로 ESG 컨설팅 사업한다.” 한국경제신문, 지면 A14면.
- 테런스 J. 세즈노스키 지음, 안진환 옮김(2018), 딥러닝레블루션, 한국경제신문.
- 하워드 솔츠, 조앤 고든(2020), 그라운드 업, 행복한 북클럽.
- Bordoloi, S., Fitzsimmons, J., Fitzsimmons, M.(2019), Service Management: Operations, Strategy, Information Technology 9th Edition, McGraw-Hill.
- Hair Jr, J. F., Blac, W. C., Babin, B. J., Anderson,

- R. E.(2019), Multivariate Data Analysis,
8th edition, Cengage.
- Oliver, R. L.(1997), Satisfaction: A Behavioral
Perspective on the Customer,
McGraw-Hill
- Ostrom, A, Iacobucci, D.(1995), Customer Trade-offs
and the Evaluation of Services, Journal
of Marketing 59, January, pp. 17-28.

[부록1]

```
import numpy as np
import pandas as pd
from scipy import stats
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.formula.api import logit
%matplotlib inline

# Model selection using Logistic Regression analysis
from sklearn import model_selection
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
clf = LogisticRegression()
df = pd.read_csv('C:/Users/gskim/Desktop/coffee/data.csv')
df = pd.DataFrame(df, columns= ['x6', 'x7', 'x8', 'x9', 'x10', 'x11', 'x12', 'x13',
'x14', 'x15', 'x16', 'x17', 'x18', 'x23'])
X = df[['x6', 'x7', 'x8', 'x9', 'x10', 'x11', 'x12', 'x13', 'x14', 'x15', 'x16', 'x17',
'x18']]
Y = df['x23']

# Logistic Analysis
affair_mod = sm.Logit.from_formula("x23 ~ x6 + x7 + x8 + x9 + x10 + x11 + x12
+ x13 + x14 + x15 + x16 + x17 + x18", df).fit()
print(affair_mod.summary())
print(np.exp(affair_mod.params))

# Model fitness
clf.fit(X, Y)

# Forecasting depend on model
y_pred = clf.predict(X)

# Model Identification
print("{} Result of Classification :
\n{}\n".format(clf, metrics.classification_report(Y, y_pred)))
print("Confusion matrix:\n{}".format(metrics.confusion_matrix(Y, y_pred)))
print("Accuracy:\n{}".format(metrics.accuracy_score(Y, y_pred)))
print("ROC_AUC:\n{}".format(metrics.roc_auc_score(Y, y_pred)))
```


[부록2]

```

from sklearn.datasets import make_classification
from sklearn.metrics import roc_curve
recall = 319 / (319 + 74)
fallout = 66 / (66 + 235)
print(recall, fallout)
fpr, tpr, thresholds = roc_curve(Y, clf.decision_function(X))
fpr, tpr, thresholds
fpr, tpr, thresholds = roc_curve(Y, clf.predict_proba(X)[:, 1])
fpr, tpr, thresholds
plt.plot(fpr, tpr, 'o-', label="Logistic Regression")
plt.plot([0, 1], [0, 1], 'k--', label="random guess")
plt.plot([fallout], [recall], 'ro', ms=10)
plt.xlabel('Fall-Out')
plt.ylabel('Recall')
plt.title('Receiver operating characteristic example')
plt.show()
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
import seaborn as sn
import matplotlib.pyplot as plt
df = pd.read_csv('C:/Users/gskim/Desktop/coffee/data.csv')
df = pd.DataFrame(df, columns= ['x6', 'x7','x8','x9', 'x10', 'x11', 'x12', 'x13',
'x14', 'x15', 'x16', 'x17', 'x18','x23'])
X = df[['x6', 'x7','x8','x9', 'x10', 'x11', 'x12', 'x13', 'x14', 'x15', 'x16', 'x17',
'x18']]
y = df['x23']

X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3,random_state=0)

logistic_regression= LogisticRegression()
logistic_regression.fit(X_train,y_train)
logistic_regression= LogisticRegression()
logistic_regression.fit(X_train,y_train)
y_pred=logistic_regression.predict(X_test)
confusion_matrix = pd.crosstab(y_test, y_pred, rownames=['Actual'],
colnames=['Predicted'])
sn.heatmap(confusion_matrix, annot=True)
print('Accuracy: ',metrics.accuracy_score(y_test, y_pred))
plt.show()

```

[부록3]

```
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
import matplotlib.pyplot as plt
seq_drop = Sequential()
seq_drop.add(Dense(10, activation='relu', input_shape=(13,)))
seq_drop.add(Dense(8, activation='relu'))
seq_drop.add(Dense(1, activation='sigmoid'))
seq_drop.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])
seq_drop.fit(X_train, y_train, epochs=100,
batch_size=32, verbose=0)
loss_drop, acc_drop = seq_drop.evaluate(X_test, y_test, verbose=0)

print('Test accuracy: %.3f' % acc_drop)
plt.show()
```

Development of customer satisfaction prediction model using deep learning

<Abstract>

The ICT-based 4th Industrial Revolution and the spread of COVID-19 are accelerating the big blur phenomenon in which manufacturing and service areas are gradually becoming ambiguous. In this situation, there is a need to clearly analyze data-based customer needs and expectations and reflect them in decision-making. In the industry, cases of applying actual machine learning and deep learning-based artificial intelligence to practice are emerging, but research on this is still insufficient in academia. This study is to develop a customer satisfaction prediction model using machine learning and deep learning for customer-related survey data of A coffee shop. It is also to present the process of developing a customer satisfaction prediction model. The logistic regression analysis technique used to predict when the dependent variable is bivariate (0, 1) is applied, and in this study, whether or not to visit the customer in the future (0, 1) is used as the dependent variable, and the model prediction performance is compared through comparison of learning data and verification data. Next, artificial intelligence analysis will be conducted using TensorFlow, an application program that is an open source software library produced by Google Brain Team and a library specializing in deep learning. As a result of the analysis, the accuracy of the customer satisfaction prediction model was 71.8%.

Keywords: big blur, artificial intelligence, machine learning, deep learning, customer satisfaction prediction model, tensorflow, accuracy.