

# 리조트 교차판매 예측모형 개발 및 SHAP을 이용한 해석

## Development of a Resort's Cross-selling Prediction Model and Its Interpretation using SHAP

강보람 · 안현철\*

국민대학교 비즈니스IT전문대학원

### 요약

관광산업은 최근 코로나19 유행으로 인해 위기에 봉착해 있으며, 이를 극복하기 위해 무엇보다 수익성 개선이 매우 중요한 상황이다. 이 때 여행 수요 자체가 축소된 코로나19와 같은 상황에서는 수익 증대를 위해 객실 점유율을 높이기 위한 공격적인 영업전략보다 어려운 여건 속에서도 찾아온 고객에게 객실 외 추가상품을 판매하여 객단가를 높이는 방향이 더 효율적일 것이다. 국내 관광 연구 분야에서 머신러닝 기법은 수요예측을 중심으로 연구된 바 있으나 교차판매 예측에 대해서는 연구된 바가 거의 없다. 또한 넓은 의미로는 호텔과 같은 숙박업종 이지만 회원제 중심으로 운영하며 숙박과 취사에 적합한 시설을 갖추고 있는 리조트 업종에 특화된 연구는 더욱이 전무한 실정이다. 이에 본 연구에서는 실제 리조트 회사의 투숙 데이터로 다양한 머신러닝 기법을 활용하여 교차판매 예측 모형을 제안하고자 한다. 또한 설명가능한 인공지능(eXplainable AI) 기법을 적용해 교차판매에 영향을 미치는 요인이 무엇인지 해석하고 어떻게 영향을 미치는지 실증 분석을 통해 확인해 보고자 한다.

■ 중심어 : 교차판매, 기계학습, 리조트, 설명가능한 인공지능, SHAP

### Abstract

The tourism industry is facing a crisis due to the recent COVID-19 pandemic, and it is vital to improving profitability to overcome it. In situations such as COVID-19, it would be more efficient to sell additional products other than guest rooms to customers who have visited to increase the unit price rather than adopting an aggressive sales strategy to increase room occupancy to increase profits. Previous tourism studies have used machine learning techniques for demand forecasting, but there have been few studies on cross-selling forecasting. Also, in a broader sense, a resort is the same accommodation industry as a hotel. However, there is no study specialized in the resort industry, which is operated based on a membership system and has facilities suitable for lodging and cooking. Therefore, in this study, we propose a cross-selling prediction model using various machine learning techniques with an actual resort company's accommodation data. In addition, by applying the explainable artificial intelligence XAI(eXplainable AI) technique, we intend to interpret what factors affect cross-selling and confirm how they affect cross-selling through empirical analysis.

■ Keyword : Cross-selling, Machine Learning, Resort, XAI, SHAP

## I. 서론

코로나 19로 인해 2020년 관광산업의 업종별 소비지출규모(1월~8월)는 전년 대비하여 여행업(↓79.6%), 관광숙박업(↓42.3%), 카지노(↓69.6%), 항공사(↓69.4%), 면세점(↓70.6%) 등 렌터카업(↑5.3%)을 제외한 대부분의 관광업종에서 하락세를 면치 못했다(한국문화관광연구원, 2020). 최근의 코로나19와 같은 전염병 뿐만 아니라 관광산업은 정치, 자연재해 등 다양한 내외부 요인에 많은 영향을 받는 산업이다(Kuo et al., 2008; 권혁진, 지윤호, 2020).

최근 관광숙박업의 위기를 극복하기 위해 무엇보다 수익성 개선이 매우 중요한 상황이다. 객실 수익의 증대 방법으로 객실 수익의 구성요소가 되는 객실수, 점유율, 평균 객실료를 최고로 하는 것이다. 나아가 객실 판매로 인한 객실 수익은 객실 수익 자체만의 수익으로 그치는 것이 아니라 호텔 영업 수익 전반에 영향을 미치는 것으로 식음료와 부대 사업장의 수익 발생은 숙박객 증대에 따른 객실 수익에 의존하고 있다(강현신, 2002).

새로운 고객을 창출하는 것은 기존 고객과 좋은 관계를 유지하여 지속적으로 구매하도록 하는 것보다 더 많은 비용이 든다는 점은 이미 선행연구자들에 의해서 지적되어 왔고 여러 각도에서 연구되어 왔다(정유경, 2003). 따라서 기존 고객과 우호적인 관계를 유지하면서, 기존 고객으로부터 새로운 매출을 유도할 수 있는 교차판매에 대한 관심이 최근 높아지고 있다.

교차판매는 기업의 기존 고객에게 새로운 상품이나 서비스를 권유하는 과정을 말한다. 상품의 특성과 고객 프로파일의 관계를 분석하여 고객들의 개별 정보, 즉 인적사항, 라이프 스타일, 구매특성 등과 매치가 되는 다른 상품들의 구입을 권유할 수 있게 된다. 상승판매는 교차판매의 한가지 형태로써 고객의 구매 형태와 관련있는

새로운 상품을 권유하는 경우를 의미한다(정유경, 2003).

여행 수요 자체가 축소된 코로나19와 같은 상황에서는 수익 증대를 위해 객실 점유율을 높이기 위한 공격적인 영업전략보다 어려운 여건 속에서도 찾아온 고객에게 객실 외 추가상품을 판매하여 객단가를 높이는 방향이 더 효율적일 것이다.

이런 가운데 타 학문 분야 및 산업 분야에서 적극적으로 도입하고 있는 머신러닝(machine learning: 기계학습) 기법은 다양한 예측 연구에서 탁월한 성과를 나타내고 있다. 컴퓨터가 데이터를 통해 모형을 개발하고 결과를 예측하는 머신러닝 기법은 타 연구 기법과 비교하였을 때, 시간과 비용은 감축시키면서 높은 예측 정확성을 보인다(이인지, 윤현식, 2020). 국내 관광 연구 분야에서 머신러닝 기법은 수요예측을 중심으로 연구된 바 있으나 교차판매 예측에 대해서는 연구된 바가 거의 없다. 또한 넓은 의미로는 호텔과 같은 숙박업종 이지만 회원제 중심으로 운영하며 숙박과 취사에 적합한 시설을 갖추고 있는 리조트 업종에 특화된 연구는 더욱이 전무한 실정이다.

이러한 배경에서 본 연구의 목적은 다음과 같이 정의할 수 있다. 실제 리조트 회사의 투숙 데이터로 다양한 머신러닝 기법을 활용하여 교차판매 예측 모형을 제안하고자 한다. 또한 설명가능한 인공지능(explainable AI, 이하 XAI) 기법을 적용해 교차판매에 영향을 미치는 요인이 무엇인지 해석하고 어떻게 영향을 미치는지 실증 분석을 통해 확인해 보고자 한다. 이를 통해 실무적으로 기업에서는 매출 증대를 위한 영업 전략 수립에 반영할 수 있을 것으로 기대 된다.

## II. 이론적 배경

### 2.1 관광 숙박업에서 애널리틱스 연구 동향

객실 상품은 성수기와 비수기 또는 주중과 주말 등 영업이 특정 계절이나 특정 요일에 편중되는 경향이 있다는 점과 일반상품과 비교하여 시간적, 공간적 제약이 있다는 특징을 지니고 있는 상품이기 때문에 호텔경영에 있어서 무엇보다도 정확하고 타당성있는 수요예측이 선행되어야 한다(김태구, 송두석, 2006).

주로 호텔 및 관광부문에 대한 수요예측은 단변량 시계열 자료 혹은 특정 사건의 개입을 고려한 이변량 시계열 자료를 이용한 분석이 많이 이용되고 있다. 김태구와 송두석(2006)은 서울 특1급 호텔의 외국인 이용객 호텔객실 수요예측모형을 제시하였는데, 계절 ARIMA모형과 개입 ARIMA모형을 적용하였다. 신현규 외(2014)는 Holt모델과 Winters 기법과 승법, 그리고 회귀모델을 통해 대전지역 호텔객실 수요예측을 위한 최적수요모델을 선정하였다.

미래에 창출될 잠재 관광객의 수요를 예측하는 관광수요 예측은 관광 상품 개발 및 정책수립 시 근거 자료로 중요하게 활용되며 관광사업에 대한 목표설정이나 타당성 검토를 위해서도 필수적이다(Jun et al., 2018).

관광수요 예측은 과거의 자료를 기반으로 지식정보시스템에서 제공하는 통계자료나 지자체에서 집계하는 통계를 기반으로 구축한다. 하지만 정확한 통계자료를 제공받기 어렵고 과거 자료에 대한 주기도 짧아지고 있으며 과거 자료가 존재하지 않는 새로운 관광지에는 적용하기 어렵다는 한계점을 제시하며 급변하는 관광환경에서 과거의 자료에만 의존하는 시계열 모형에서 벗어나 온라인 리뷰를 활용하여 감성값을 계산하여 ARIMA와 BPN(Back Propagation Neural Network), SVR(Support Vector Regression), 딥러닝 기법인 RNN(Recurrent Neural Networks)을 적

용하였으며 딥러닝 모형을 제시하기도 하였다(김은미, 2021).

수요예측 모형 외에는 고객 행동에 대한 예측으로서 에어비앤비의 고객 기본 정보 및 웹 로그 데이터를 활용하여 Lasso, SVM, 랜덤포레스트 등 다양한 방법론을 활용하여 최적의 모형을 찾고 웹 로그데이터의 유무에 따른 예측오차를 비교하여 웹 로그 데이터를 활용하여 예측한 경우 사용하지 않는 경우보다 예측의 정확도가 최대 2배 더 높아진 것을 확인하여 웹 로그 데이터의 효용성을 확인하기도 하였다(안효인 외, 2019).

### 2.2 여행업에서 교차판매 연구

호텔 업종에서는 2000년대 초반에 객실 룸 업그레이드를 위한 제안, 식음업장에서 세트메뉴 추천 등을 하는 상승판매 기법이나 활성화 방안 등에 대한 연구가 있었다. 강현신(2002)은 호텔의 높은 수익과 점유율, 고객만족을 위한 판매기술로서 추천판매가 활성화 되기 위해 직원들의 인식 제고, 주변 호텔과의 실무적인 협의, 예약업무 매뉴얼의 개발, 직원에 대한 보상 등의 방법을 제시하였다. 고객에게 호텔 객실에 대해 안내할 때 직원이 다양한 상품과 서비스를 제안하여 정보제공 역할을 함으로써 고객이 좀 더 훌륭한 객실, 서비스, 시설 등을 선택했을 때에는 호텔에서 즐겁게 머물게 되며 고객의 만족을 창출하기 위하여 호텔직원이 추천판매방법을 사용하여야 하는데, 추천판매는 고객을 가능한 한 고가의 객실에 머물도록 하는 판매기술이 필요하며 추천판매는 호텔의 매출을 신장시키는데에 중요한 역할을 할 뿐만아니라 다른 부대시설의 판매촉진을 통해 매출신장에 기여할 수 있다고 하였다.

전홍진(2007)은 객실상품과 식음료 상품에 대해서 상승판매의 개념과 접근방법을 제시하고 있다. 판매 단계별 구체적인 제안 방법으로 제안 판매나 특정 품목을 추천, 메뉴 설명, 대안 제시 등

의 실무적인 기법을 제안하기도 하였다.

정유경(2003)은 연관규칙을 활용해서 호텔 고객의 데이터베이스를 이용하여 로비 라운지, 중식당, 양식당, 베이커리 카페 등 이용간에 지지도와 신뢰도, 리프트값을 통해 어떤 연관성이 존재하는지 파악하는 연구를 하였는데, 자료로부터 발견된 연관 규칙을 통해 리프트가 높게 나온 고객들에게 교차판매 또는 상승판매를 제안하고, 지지도와 리프트가 높은 고객에게는 교차판매 캠페인을 진행하는 것을 제안하였다.

### 2.3 교차판매를 애널리틱스로 접근한 기타 연구

교차판매는 금융, 보험, 통신 분야에서 머신러닝 기법을 활용하여 활발하게 연구되어 왔다. 진서훈과 최종훈(2008)은 캐피탈업에서의 교차판매 모델에 대한 연구로 데이터마이닝 기법을 활용하여 교차판매 캠페인 반응 스코어모델을 개발하였다. 기존 고객 정보를 활용하여 진행되는 마케팅 활동으로 타겟 마케팅의 형태로 진행될 때 비용 및 효율성의 이유로 기업이 보유한 모든 고객에게 진행할 수 없으므로 교차판매 오피에 반응률이 높을 것으로 예상되는 고객을 선별하기 위해 스코어 모델을 활용할 것을 제안하였는데, 해당 모델은 스코어링 점수를 도출하는 것이 중요하기 때문에 로지스틱 회귀 모형 하나만을 가지고 각 변수에 대한 회귀계수를 활용하였다.

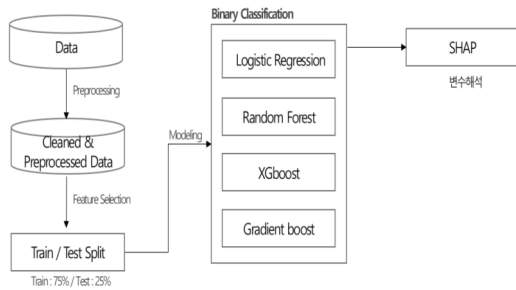
한상태 외(2004)의 연구 또한 교차판매 캠페인에 활용할 수 있도록 국내 손해보험사의 고객 데이터베이스를 통해 데이터마이닝 모형을 만들었다. 자동차보험에 가입한 보험사 고객 중에서 장기보험 및 세부보험에 추가로 가입하는 고객의 특성을 파악하기 위해 스코어링 모형을 개발하였다. 모형을 근거로 캠페인 활동을 통해 마케팅의 효율성을 증대시키는데 활용하는 것을 제안하였다. 모형은 의사결정나무, 로지스틱 회귀분석, 신

경망 분석 세 가지를 활용하였으며 로지스틱 회귀분석이 가장 높은 정확도 83%를 보였다. 이에 로지스틱 회귀분석의 결과를 활용하여 각 보험상품별로 가입확률을 도출하여 가입확률이 높을 것으로 예상되는 상품을 추천하는 캠페인 활동을 제안하였다.

Ahn et al.(2011)은 이동통신시장에서 교차판매를 일으키기 위한 새로운 모형을 제안하였다. 기존의 조합 방법은 가중평균 방법을 적용하여 예측하지만 두 개의 임계값을 사용하여 최종결정에 효과적으로 점수를 매기게 되어 예측정확도를 높이는 모델을 만들었다. 여러 요인을 동시에 최적화 하기 위해 유전자 알고리즘(GA)을 사용하여 가중치와 임계값을 최적화 하는 도구로 사용하였다.

## III. 연구모델

기존의 교차판매 모델에서는 고객의 나이, 성별, 연령대 등 기본적인 개인정보를 활용하였으나 최근 개인정보 보호법이 강화되고 유출에 대한 사고로 인해 관련한 이슈를 기업에서는 더욱 엄격하게 다루고 있으며, 특히 리조트는 회원권을 보유한 개인 또는 법인이 예약하고 실제 투숙자는 개인회원의 추천인이나 법인회원의 임직원으로써 예약자와 투숙자가 상이한 경우가 많다. 따라서 방문 전에 투숙자의 정보를 알기 어렵고 예약자의 개인정보를 기반으로 교차판매 캠페인을 실행하기 적절하지 않을 수 있다. 이에 본 연구에서는 투숙자의 추가적인 개인정보 없이 상품 정보 만으로도 활용할 수 있는 교차판매 예측 모델을 제안한다. 이는 캠페인으로 활용할 수도 있겠지만 XAI 기법을 활용하여 모델을 해석함으로써 객실 외 추가적인 매출을 일으키는 영업전략을 수립하는데에 실무적인 시사점을 줄 수 있을 것으로 기대한다. 본 연구에서 진행되는 실험의 전반적인 구조는 <그림 1>과 같다.



〈그림 1〉 연구모델

수집된 원본 데이터를 전처리한 후 데이터를 학습용과 검증용 데이터의 비율은 75% : 25%로 분류한 다음 학습용 데이터는 과적합을 방지하기 위해 중속변수인 교차판매 여부를 1:1 비율이 되도록 오버 샘플링을 하였다. 그 다음 로지스틱 회귀분석 (Logistic Regression), 랜덤포레스트 (RandomForest), XGBoost, Gradient Boost 네가지 기법을 사용하여 예측률을 확인하고 그 중 우수한 모델을 선택하여 XAI 기법 중 하나인 SHAP(SHapley Additive exPlanations)를 활용하여 변수들의 중요도를 해석하고자 하였다.

SHAP는 샐플리 값과 피쳐 간 독립성을 핵심 아이디어로 사용하는 XAI 기법으로 샐플리 값을 통해 중속변수 결과에 어떤 독립 변수들이 기여를 했는지 수치로 표현할 수 있다. 샐플리값은 중속변수의 기여도는 그 중속변수의 기여도를 제외했을때에 중속변수 결과의 변화 정도를 나타내어 추론하는 방법으로 모델의 출력을 각 변수들의 기여도로 분해할 수 있다.

한 개의 트리를 사용하는 의사결정 나무 외에 트리 모델을 여러개 사용, 결합하여 성능을 높이는 방법으로 앙상블 모델이 있다. 여러 모델의 예측을 종합하면 모델의 예측력은 일반적으로 좋아 지는데, 본 연구에서는 배깅(bagging)의 방법 중 하나인 랜덤포레스트와 부스팅(boosting) 방법인 XGBoost, Gradient Boost를 사용한다. 배깅은 같은 종류의 알고리즘 모델을 여러 개 학습하여 예

측하는 것이며, 부스팅은 여러개의 약한 학습기를 순차적으로 학습하여 잘못 예측한 데이터에 대한 예측오차를 줄일 수 있는 방향으로 모델을 계속 업데이트 하는데, 여러 모델을 동시에 학습하지 않고 순서대로 학습하는 점에서 배깅과 다르다(오승환, 2021).

## IV. 실증분석

### 4.1 실험 데이터

본 연구에서는 국내 H리조트 체인의 2021년 1월~12월 기간의 1년치 투숙 데이터를 제공받아 분석에 사용하였다. 해당 데이터에는 투숙객의 고객 개인정보는 포함되어 있지 않으며, 객실 상품에 대한 정보와 예약에 관련한 속성만이 포함되어 있다. 분석에 사용된 데이터는 총 484,857건이다.

중속변수값은 객실당 결제 정보의 교차판매 여부로 교차판매는 객실 외 리조트 내 식음시설이나 테마시설 등에 지출한 경우를 의미하여 객실 외 상품에 지출한 경우를 1, 객실만 이용한 경우를 0으로 값을 부여하였다. 중속변수를 예측하기 위해 활용되는 독립변수는 총 74개 중 독립표본 t검정과 카이제곱 검정을 통해 통계적으로 유의한 변수로 65개를 최종적으로 선정하였다. <표 1>은 본 연구에서 사용된 독립변수의 목록이다.

투숙한 일자에 따라 성수기, 방문 월, 주말, 요일 특성을 추출하였으며 머무르는 기간과 예약한 객실수에 따라 박수, 룸수 변수를 추출하였고, 어떤 지역의 지점을 이용하였는지와 예약부터 방문까지의 소요일자를 나타내는 리드타임을 연속형 변수와 함께 당일, 전일, 3일, 일주일 등 유의미한 구간으로 범주화하여 변수로 활용하였다. 그 외에 객실타임, 결제방법, 예약채널, 예약유형 등 주로 범주형 변수가 모델에 활용되었다.

<표 1> 실험에 사용된 변수

구분	변수명	설명
상품 특성	summer_yn	성수기여부
	month1~month 12	방문 월
	weekend_yn	주말 여부
	sunday~saturday	방문 요일
	nights	박수
	rooms	룸수
	continue_yn	연박여부
	manynights_yn	다실여부
	roomonly_pkg_yn	특가패키지 여부
지불 구분	branch	9개 지점
	pay_div	지불구분_분납
	pay_prepay_online	지불구분_선결제
	pay_pre	지불구분_선납
	pay_on_site	지불구분_현지불
예약 채널	pay_after	지불구분_후불
	channel_callcenter	예약채널_고객센터
	channel_ARS	예약채널_ARS
	channel_staff	예약채널_본사(직원)
	channel_branch	예약채널_사업장
	channel_pc	예약채널_pc
예약 일- 투숙일	channel_mobile	예약채널_모바일
	channel_channelmanager	예약채널_채널매니저
	lead_time_1month	1개월 이내
	lead_time_2month	2개월 이내
	lead_time_3month	3개월 이내
	lead_time_3week	3주 이내
	lead_time_2week	2주 이내
	lead_time_7day	7일 이내
객실 타입	lead_time_3day	3일
	lead_time_yesterday	전일
	lead_time_today	당일
예약 유형	hotel_type_yn	호텔형 객실타입
	kids_type_yn	키즈형 객실타입
	upgrade_room	업그레이드 객실타입
	bookingtype_ota	예약유형_OTA
	bookingtype_mem	예약유형_기명(개인회원)
	bookingtype_cor	예약유형_기명(법인회원)
	bookingtype_mem_any	예약유형_무기명(개인회원)
	bookingtype_cor_any	예약유형_무기명(법인회원)
	bookingtype_coupon	예약유형_선납권/초대권
	bookingtype_cor_staff	예약유형_법인회원_임직원
bookingtype_mem_online	예약유형_온라인회원	
bookingtype_discor	예약유형_할인사일반	
bookingtype_mem_coupon	예약유형_회원 주중 선납쿠폰	

### 4.2 적용된 분류모델

본 연구에서는 전통적인 분류기법인 로지스틱 회귀분석, 트리 모델의 종류인 랜덤포레스트, XGBoost, Gradient Boost까지 총 4개 기법의 비교 모형을 설정하고 예측성능을 비교해 보았다. 각 모형별 변수와 데이터셋은 동일하며 모두 파이썬 Scikit-learn 라이브러리를 활용하여 분석하였다.

각 모형의 실험설정은 랜덤포레스트는 모형 (weak learner)의 개수는 50개(n\_estimators=50), 불순도 지정은 지니(criterion='gini'), 트리의 최대 깊이는 4(max\_depth=4)로 설정하였다. XGBoost에 사용한 부스터 구조는 gbtree, 트리의 최대 깊이는 3이며 GradientBoost의 불순도 지정은 friedman\_mse, 트리의 최대 깊이 역시 3으로 설정하였다.

### 4.3 실험결과 해석

총 네 개의 기법을 활용하여 실험을 한 결과 <표 2>에 제시되어 있는 바와 같이 가장 예측정확도를 보이는 기법은 XGBoost였다. 또한 각 모델의 예측정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score (Precision과 Recall의 조화평균) 등을 비교하여 전반적인 성능 향상이 있는지를 확인하였다. <표 3>에 제시되어 있는 바와 같이, 검증 데이터에 대하여 성능을 비교해보면 Precision은 네 개의 모델이 비슷한 결과를 보였고 전체 데이터에서 실제로 1인 경우 중 모델이 1이라고 정확하게 예측한 비율인 Recall이 가장 XGboost에서 높게 나타났으며, 종합적으로 판

<표 2> 모형간 정확도

정확도	Logistic Regression	Random Forest	XGBoost	Gradient Boost
학습 데이터	0.68	0.63	0.68	0.68
검증 데이터	0.71	0.71	0.73	0.71

〈표 3〉 검증 데이터 기준 모형간 성능 비교

정확도	Logistic Regression	Random Forest	XGBoost	Gradient Boost
Precision	0.81	0.79	0.81	0.81
Recall	0.71	0.71	0.73	0.72
F1-score	0.74	0.74	0.76	0.75

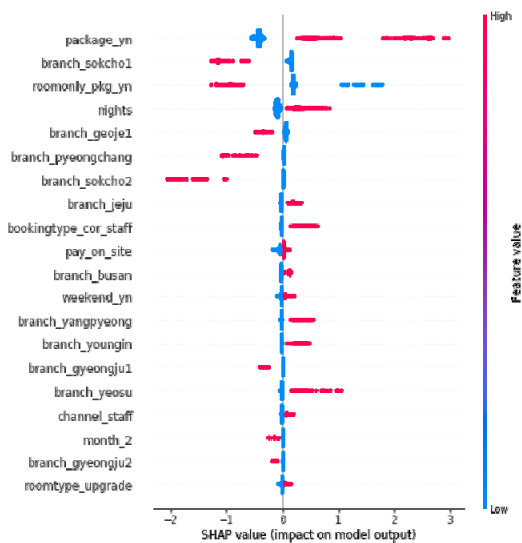
단할수 있는 지표인 F1-score도 XGboost가 0.76으로 우수하게 나타났다.

#### 4.4 XAI 적용 결과

블랙박스 모델인 XGBoost에 SHAP 기법을 적용하여 어떤 변수가 어떤 방향으로 교차판매 예측에 영향력을 미치는지 확인해 보았다. <그림 2>는 SHAP 기법을 적용하여 도출한 전역적인 변수 영향도 그래프이다. SHAP value에 나타난 각 점들의 색은 종속변수에 대한 영향도가 클수록 붉은색을 띠고, 영향도가 작을수록 파란색을 보인다. 붉은색 점은 해당변수가 종속변수를 결정하는 데에 큰 영향을 미쳤음을 의미하고, 파란색 점은 변수가 종속변수를 결정하는데 적은 영향을 미쳤음을 의미한다. 그리고 각 점들이 그래

프에서 X축을 기준으로 음의 값을 가지면 결과에 부(-)의 영향을, 양의 값을 가지면 정(+)의 영향을 미친다는 것을 의미한다.

<그림 2>는 변수별 SHAP 값을 계산하고 영향도를 점으로 시각화한 그래프로 영향도가 높은 20개 변수에 대해 순서대로 그래프로 표현되었다. 영향도를 확인하기 위해 푸른색 대비 붉은색이 더 많이 분포된 변수 위주로 살펴보고, 붉은색 점의 방향으로 해석해 보았다. 주로 상품특성, 지점, 예약유형 중 법인회원 임직원인 경우 교차판매에 영향을 미치는 것으로 나타났다. X축을 기점으로 오른쪽에 있는 교차판매에 정(+)의 영향을 미치는 변수는 위에서부터 순서대로 패키지 구매 여부 (package\_yn), 연박 여부(nights), 제주 (branch\_jeju), 양평(branch\_yangpeong), 용인 (branch\_youngin), 여수(branch\_yeosu), 지점 변수와 법인회원의 임직원 (bookingtype\_cor\_staff) 변수이다. 교차판매에 부(-)의 영향을 미치는 변수는 위에서부터 순서대로 지점 변수인 속초 (branch\_sokcho1), 객실 임박할인 상품 구매여부 (roomonly\_pkg\_yn), 거제(branch\_geoje1), 평창 (branch\_pyeongchang), 경주 (branch\_gyeongju1)로 나타났다.



〈그림 2〉 SHAP 기법을 적용한 변수 영향도

패키지는 객실과 식음, 테마시설 등 타 유형의 부대상품과 결합하여 판매하는데, 리조트 특성상 부부와 자녀를 동반한 4명 이상의 가족단위의 고객들이 주로 방문한다. 따라서 객실 인원도 4인 기준이 대부분인데, 부대상품은 2인 기준으로 상품이 만들어진다. 패키지 상품을 구매한다는 것은 이미 리조트 내에 다른 시설들을 이용하려는 의도를 가지고 방문한 것이며, 추가 인원에 대한 지불로써 교차판매가 발생하게 된다.

반면 상품 유형 중에서 객실 임박할인 상품 구매할 경우 교차 판매에 부의 영향을 미쳤다. 저렴한 가격의 상품으로 저렴한 예산으로 임박해서 떠나는 경우, 즉 특가 상품으로 예약하여 이용하는 경우에 리조트 내 추가적인 상품을 구매하는

경우가 적었다. 임박하여 떠나는 즉흥 여행의 경우 동반인원수가 적을 것이고 특정 장소를 목적으로 떠나는 목적성 여행인 경우가 많아 숙소 내부 시설에 대한 이용 니즈가 적은 것으로 해석해 볼 수 있다.

또한 2박 이상 숙소에 머무르는 경우 교차판매를 하는 경우가 높았다. 1박 2일로 짧게 여행을 한다면 근처 관광지나 맛집을 둘러보다가 오후 늦게 입실하여 다음날 오전에 퇴실을 하게 되는데 이 경우 숙소 내 다른 서비스를 경험할 가능성이 낮고 2박이상 머무르는 경우 여유로운 일정으로 숙소 내에 상품들과 서비스를 경험하는 경우가 많다고 생각해볼 수 있다.

객실 예약 유형 중 법인회원의 임직원인 경우, H리조트에서는 별도의 재직 인증을 하게 되면 법인회원권을 관리하는 담당자를 거치지 않더라도 회원가적으로 예약을 할 수 있는데, 번거로울 수 있는 재직 인증절차를 거친 법인회원의 임직원인 경우 H리조트 체인의 시설들을 주기적으로 이용할 의지가 있는 고객군으로 볼 수 있기에 객실 외 다양한 상품을 이용하는 교차판매를 일으킬 가능성이 높다고 할 수 있다.

그 외에 방문 지점별 교차판매 여부에 대한 정, 부의 영향을 살펴보면 국내 여행지 중 체류기간이 길어 2박이상 연박을 하는 경우가 많은 제주 지점에 머무르는 경우 교차판매를 일으킬 가능성이 높고, H리조트의 양평, 용인, 여수 지점은 직영 식음시설에서 객실 내 배달 서비스 등 식음 시설에 대한 경쟁력이 우수하여 교차판매가 잘 일어난 지점이라고 볼 수 있다. 반면 속초, 거제, 평창, 경주 지점을 이용하는 경우 교차판매에 부의 영향을 미쳤는데, 위 지역의 경우 관광지로서 외부에 이용할만한 대체 식음시설들이 많아 늦은 입실과 늦은 퇴실을 하게 되어 리조트 내 교차판매가 적다고 볼 수 있다.

## V. 결 론

### 5.1 연구의 요약

자연재해, 전염병 등 대내외적 변수들에 취약한 관광산업, 특히 숙박업에서 이러한 위기를 극복하기 위한 수익성 개선의 방법으로 교차판매 전략을 수립할 수 있다. 신규 방문을 위한 영업에 자원을 투여하는 것 보다 방문을 한 고객의 객단가를 높이는 방향이 더 효과적일 것이라고 생각한다. 교차판매 관련하여 통신, 보험, 금융등 산업에서는 교차판매 캠페인 전략이라던지 어떤 요인이 영향을 미치는지 규명하는 연구가 진행되었으나, 관광 숙박업에서는 거의 연구가 이루어지지 않았다.

본 연구에서는 국내 H리조트 체인의 투숙데이터를 활용하여 교차판매를 예측하고, 예측성능을 비교해 보았다. 총 74개 변수 중 카이제곱 검정과 T-test를 통해 유의한 변수 65개를 최종 선정하여 교차판매 여부를 예측하는 이분류 모델을 로지스틱 회귀분석과 트리 모델 중 배깅과 부스팅 기법을 사용한 랜덤 포레스트, XGBoost, Gradient Boost 총 4개의 분류모델을 만들었다.

가장 뛰어난 예측 정확도를 보인 모델은 XGBoost 모형으로 73% 예측 정확도를 확인하였다. 또한 XGBoost에 SHAP기법을 적용하여 어떤 변인들이 교차판매에 영향을 미치는지 해석하였다. 주로 특가 임박 상품으로 즉흥적이고 주변 관광지에 목적을 두어 단순 숙박을 위한 방문 보다는 예약 단계에서 리조트의 객실 외 부대시설 상품들을 구매 고려하고 인지하여 패키지 상품으로 구매하거나, 체류시간이 긴 경우, 리조트 주변보다 리조트 내 식음업장의 메뉴 경쟁력이 높은 경우 교차판매를 일으켜 보다 높은 객단가를 지출하여 매출 증대에 도움을 줄 수 있다고 해석하였다.

이런 해석으로 인하여 온라인에서 객실 판매시 예약단계에서 부대 시설에 대한 안내나 묶음 상품을 우선 노출 시키고, 1박보다 연박 상품에



서 긴 체류시간 안에서 즐길만한 서비스들을 안내 하는 전략을 수립할 수 있을 것이다. 또한 방문한 지점별로 교차판매 여부가 상이하어 교차판매가 우수한 지점을 벤치마킹하여 타 지점의 영업전략에 반영할 수 있을 것이다. 이렇듯 변수들의 해석으로 교차판매를 촉진시키기 위한 전략을 세울 때 이런 지식들이 실무적으로 유용하게 활용될 수 있을 것으로 기대된다.

## 5.2 연구의 의의

본 연구는 기존에 호텔 관광분야에서 잘 다루지 않았던 교차판매를 주제로 머신러닝을 활용한 예측모형을 제안했고 모델만 구축한 것이 아니라 결과를 해석을 시도했다는 점에서 학술적인 기여가 있다. 또한 제안 모델에 사용된 데이터는 고객 개인정보가 포함되어 있지 않다. 투숙자의 인구통계학적 특성이나 개개인의 정보를 포함하지 않았음에도 73%의 예측률을 보여 고객 개인정보 유출에 대한 위험 없이도 교차판매를 예측하는 모델을 제안하였다. 따라서 민감한 정보를 포함하지 않고 예약전 단위로 실무에서 교차판매를 유도할 수 있도록 객실 단위의 마케팅을 하는데 활용할 수 있을 것이다.

## 5.3 연구의 한계점 및 향후 연구방향

본 연구에서 사용된 데이터는 개인정보를 사용하지 않았지만 재방문 여부, 동반자 유형, 연령대에 따라 여행 성향이 상이한 만큼 방문자에 대한 개인의 정보가 포함되었다면 예측률은 더욱 우수했을 것이다. 또한 해당 데이터가 H리조트 체인 한회사의 데이터여서 회사 고유의 특성이 반영되었을 것으로 예상된다. 후속 연구는 투숙자의 정보와 다른 리조트의 정보를 활용한다면 더욱 우수한 예측과 풍부한 해석이 될 수 있을 것이다.

## 참 고 문 헌

- [1] 강현신, “객실수의 극대화를 위한 추천판매 활성화 방안”, 문화관광연구, 제4권, 제3호, pp.1-15, 2002.
- [2] 권혁진, 지운호, “코로나19에 따른 호텔고객의 전환의도 결정요인에 관한 연구”, 호텔리조트연구, 제19권, 제6호, pp.25-45, 2020.
- [3] 김은미, “온라인 리뷰의 감성분석과 순환신경망을 적용한 국내 인바운드 관광수요 예측 모형”, 관광연구저널, 제35권, 제3호, pp.69-79, 2021.
- [4] 김태구, 송두석, “ARIMA모형을 적용한 외국인 이용객 호텔객실 수요예측모형 선정: 서울 특1급 호텔을 중심으로”, 호텔경영학연구, 제15권, 제5호, pp.97-118, 2006.
- [5] 신현규, 김흥렬, 이준재, “대전지역 호텔객실에 대한 시계열 예측방법에 관한 연구”, 관광연구저널, 제28권, 제7호, pp.49-58, 2014.
- [6] 안효인, 최유리, 오래은, 송종우, “에어비앤비(Airbnb) 웹 로그 데이터를 이용한 고객 행동 예측”, 응용통계연구, 제32권, 제3호, pp.391-404, 2019.
- [7] 오승환, *파이썬 딥러닝 머신러닝 입문*, 정보문화사, 2021.
- [8] 이인지, 윤현식, “머신러닝을 활용한 지역축제 방문객 수 예측모형 개발”, 정보시스템연구, 제29권, 제3호, pp.35-52, 2020.
- [9] 전홍진, “호텔상품의 Up-selling 기법에 관한 연구”, 관광정보연구, pp.26-29, 2007.
- [10] 정유경, “연관규칙을 이용한 호텔 레스토랑 고객의 이용행태 분석”, 호텔경영학연구, 제12권, 제2호, pp.1-22, 2003.
- [11] 진서훈, 최종후, “효과적 교차판매를 위한 데이터마이닝 활용방안”, Journal of The Korean Data Analysis Society, 제10권, 제5호, pp.2629-2638, 2008.

- [12] 한상태, 강현철, 이성건, 정요천, “교차판매 (Cross-Sell) 스코어링 모형 개발 Case Study Papers : A Development of Cross-Sell Scoring Model”, 응용통계연구, 제17권, 제2호, pp.229-238, 2004.
- [13] Ahn, H., Ahn, J. J., Oh, K. J., & Kim, D. H., “Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques,” Expert Systems with Applications, vol.38, no.5, pp.5005-5012, 2011.
- [14] Jun, W., Yuyan, L., Lingyu, T., & Peng, G, “Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand forecasting,” Chaos, Solitons & Fractals, vol.108, pp.136-147, 2018.
- [15] Kuo, H.I., C.C. Chen, W.C. Tseng, L.F. Ju, & B.W. Huang, “Assessing impacts of SARS and Avian Flu on international tourism demand to Asia,” Tourism Management, vol.29. no.5, pp.917-928, 2008.

## 저 자 소 개



### 강 보 략(Boram Kang)

- 2014년 : 경상국립대학교 경영정보학과 (학사)
- 2022년 : 국민대학교 비즈니스 IT전문대학원 (석사)
- 관심분야 : 데이터 마이닝, 머신러닝, CRM



### 안 현 철(Hyunchul Ahn)

- 2006년 : KAIST 테크노경영대학원 경영공학(박사)
- 2009년~현재 : 국민대학교 비즈니스IT전문대학원 교수
- 관심분야 : 비즈니스 애널리틱스, 지능형 의사결정지원시스템