

소비데이터 기반 여행 성향 예측 모델 설계에 관한 연구 :트래블소나(Travel-Sona)를 중심으로

황미녕, 강남규, 박송연*

한국과학기술정보연구원, *(주)데이콘인피니티

{mnhwang, ngkang}@kisti.re.kr, *songyoun@dacon.kr

A Study on the Design of a Travel Preference Prediction Model Based on Consumer Data : A Personalized Approach Centered on the Travel-Sona Framework

Mi-Nyeong Hwang, NamGyu Kang, SongYoun Park*

KISTI, *Dacon Infinity

요약

카드 소비 데이터는 개인의 행동 특성을 드러낼 수 있는 대표적인 비정형 데이터로, 이를 기반으로 한 분석은 개인 맞춤형 서비스 구현을 위한 유의미한 자료가 될 수 있다. 본 연구는 카드 소비 데이터를 활용하여 개인의 소비 성향을 유형화하고, 이를 토대로 여행 성향을 예측하는 모델을 개발하는 데 목적이 있다. MBTI 성격유형 이론의 네 가지 차원(E/I, S/N, T/F, J/P)을 소비 행동 지표로 재구성하였으며, K-means 군집화, 주성분 분석(PCA), 로지스틱 회귀 등의 기법을 적용하였다. 소비 성향의 조합을 통해 정의된 '트래블소나(Travel-Sona)' 모델은 약 1,200만 건의 카드 거래 데이터를 기반으로 구축되었고, 군집 간 통계적 유의성과 예측 모델의 신뢰도에 대한 검증을 함께 수행하였다. 본 연구는 소비 데이터를 활용한 성향 분석이 관광 분야에서의 개인화된 콘텐츠 추천, 지역 기반 서비스 설계 등 다양한 영역에 활용될 수 있음을 보여준다. 특히, 데이터 기반 성향 예측 모델을 통해 향후 마이데이터 기반 서비스 기획, 공공 맞춤형 정책 수립 등에도 적용 가능성을 제시한다는 점에서 의의가 있다.

1. 서론

디지털 기술의 발달로 일상 생활의 소비 활동에서도 생성되는 데이터가 기하급수적으로 증가하고 있다. 이 가운데 신용카드의 결제/승인 데이터는 소비자 개인의 취향이 반영되어 있어 성향 분석에 활용되기에 적합하다. 최근 몇 년 사이 이 데이터를 활용해 소비자의 행동을 분석하려는 노력이 늘고 있다[1].

과거에는 소비자 분석이 주로 성별, 연령, 지역 같은 인구통계학적인 기준에 의존해왔다면 이제는 사람마다 다른 성격이나 사고방식, 행동패턴, 소비 습관까지 고려하여 분석하고자하는 경향을 볼 수 있다. 성격 유형 이론 중에서도 비교적 대중적으로 알려진 MBTI를 소비 분석에 적용해보려는 노력도 있어왔다[2].

본 연구에서는 MBTI 이론을 기반으로 카드 소비 데이터를 분석해 개인의 소비 성향을 분류하고, 이를 통해 여행 성향까지 예측할 수 있는 모델을 구성하고자 했다. 핵심은 단순한 유형 분류가 아니라, 소비자의 실제 사용 데이터 속에서 성향과 행동 간의 연관성을 찾아내는 것이며, 이를 통해 개인 맞춤형 여행 서비스 설계에도 실질적인 근거가 될 수 있는 분석틀을 마련하는 데 있다.

II. 본론

본 연구에서는 특정 카드사의 매출전표(승인/취소) 데이터를 활용하여 개인의 소비성향을 분석하였다. 그 과정은 <그림 1>과 같이 데이터 획득, 데이터 전처리, 모델링, 모델 검증 및 평가로 진행되었다.

1. 데이터셋 구성 및 전처리

실험에 사용된 카드 결제 데이터는 2023년 7월부터 2024년 6월까지 12개

월간의 데이터를 사용하였으며, 2024년 6월까지 12개월간의 카드거래 데이터를 기반으로 하였으며, 총 28,938명의 카드 결제 데이터(약 1,200만 건)를 대상으로 하였다. 본 연구에 사용된 카드 거래 데이터는 하나카드 내부의 분석 전용 시스템내에서 가명 처리된 상태로 분석되었다. 모든 분석은 하나카드사가 지정한 통제된 환경에서만 수행되었으며, 연구 종료 후에는 샘플링된 데이터 뿐만 아니라 분석 과정에서 생성된 2차 데이터 및 가공데이터까지 하나카드사의 내부 규정에 따라 폐기되는 절차를 거쳤다.

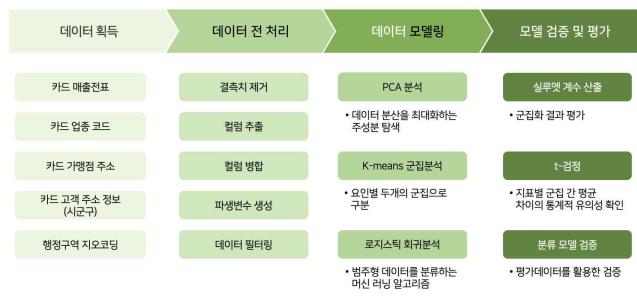


그림 1 소비성향 분석 과정의 전체 프로세스

연구 전 과정은 개인정보 보호법, 정보통신망법 및 관련 정보보호 지침을 준수하여 수행되었으며, 개인 식별이 불가능한 수준의 데이터 처리 절차가 철저히 확보되었다. 전처리 과정에서는 총 결제건수 120건 미만이거나 연간 결제금액 360만 원 미만인 데이터를 제외하여 26,846명을 최종 분석 대상으로 선정하였다. 주요 사용 데이터는 매출전표, 가맹점 주소, 고객 주소, 업종 코드 등 5종의 텍스트 데이터로 구성되었고, <그림 2>와 같이 총 13개의 주요 컬럼이 소비행태 분석에 사용되었다.

데이터명	형태	내용	출처	기온연월
카드사 매출전표	CSV	매출전표 통합 데이터의 145개 컬럼 중 33개 추출 후 13개 컬럼 사용	카드사	2024.09
카드사 업종 코드	xlsx	카드사 업종 대,중,소분류 코드 및 업종명	카드사	2024.09
카드사 가맹점 주소	CSV	카드 가맹점 주소(광역시도, 시군구, 읍면동)	카드사	2024.09
카드사 고객 주소	CSV	가맹처리된 고객의 주소(광역시도, 시군구)	카드사	2024.09
행정구역 지오코딩	CSV	행정구역별(시군구, 읍면동) 위도 및 경도	구글맵	2024.10

NO	컬럼 명	NO	컬럼 명
1	발급고객번호	9	승인번호
2	카드종류코드	10	승인일자
3	할인금액	11	승인요일구분코드
4	이용포인트	12	승인시각
5	취소여부	13	승인금액
6	일시불할부구분코드		
7	가맹점한글명		
8	당사업종코드		

그림 2 소비행태 분석에 활용된 주요 컬럼 목록

2. 소비성향 요인 정의 및 지표 설정

소비자 성향은 MBTI의 네 가지 요인(E/I, S/N, T/F, J/P)을 기반으로 구성하였다. 각 요인은 행동 지표로 전환하여 객관적 수치를 통해 성향을 구분할 수 있도록 설계하였으며, 매핑 기준은 <표 1>에서 확인할 수 있다.

표 1 트래블소나 지표

트래블소나요인	(분석지표) 카드 소비데이터 사용 컬럼
여행장소 E(외향적)/I(내향적)	·(업종)당사업종코드 ·(가맹점및거주지간거리) 가맹점 주소, 거주지 주소 ·(결제빈도) 발급고객번호, 승인일자 ·(총결제금액) 발급고객번호, 승인금액 ·(저녁및주말결제건수) 발급고객번호, 승인시각, 승인요일구분코드, 카드종류코드
여행방식 S(감각적)/N(직관적)	·(이용업종수)발급고객번호, 당사업종코드, 카드종류코드 ·(이용가맹점수)발급고객번호,가맹점한글명 ·(가맹점최대방문수)발급고객번호, 가맹점한글명, 카드종류코드 ·(온라인결제비율)발급고객번호,당사업종코드
소비 우선순위 T(사고적)/F(감각적)	·(업종)당사업종코드 ·(할인금액비율)발급고객번호, 할인금액, 승인번호 ·(할인건수비율)발급고객번호, 할인금액, 승인번호 ·(이용포인트비율) 발급고객번호, 이용포인트, 승인금액 ·(포인트이용건수비율) 발급고객번호, 이용포인트, 승인번호
소비 가치기준 J(판단적)/P(인식적)	·(업종) 당사업종코드 ·(결제취소율) 승인건수, 취소여부 ·(할부이용률)승인건수, 일시불할부구분코드 ·(월별결제건수표준편차) 승인번호, 승인일자

3. 소비성향 군집화 및 분류모델 구축

본 연구에서는 소비 성향을 설명하기 위한 네 가지 심리 요인(E/I, S/N, T/F, J/P)에 대해 PCA(Principal Component Analysis.) 방법으로 차원을 축소하고, K-means를 이용하여 각 요인별로 클러스터를 2개가 나오도록 군집화를 진행하였다. 군집화 결과는 실루엣 계수(Silhouette Coefficient)를 이용해 타당성을 검토했으며, 모든 요인에서 기준값인 0.3을 초과하는 계수를 도출하여 군집이 합리적으로 이루어졌음을 증명하였다.

각 군집에 대해서는 주요 지표의 평균값을 비교, 분석하여 성향 특성을 해석하고, 이를 바탕으로 E/I, S/N 등 해당 요인에 대한 레이블을 부여하였다. 군집 결과를 기반으로 로지스틱 회귀분석(Logistic Regression)을 수행하여 요인별 분류모델을 구축하였고, 전체 데이터는 80%를 학습용, 20%를 검증용으로 분할하여 분석 및 성능 평가를 진행하였다.

모델 성능 분석 결과, E/I, S/N, J/P 요인은 정확도 99% 이상, AUC 1.00으로 매우 높은 분류 성능을 보였다. T/F 요인의 경우 정확도는 87.85%, AUC는 0.95 수준으로 양호했으나, 정밀도(86.79%)와 재현율(97.84%) 간 차이가 크게 나타나, 사고형(T) 성향에 대한 예측에서는 일부 오분류 가능성이 존재함을 확인할 수 있었다.

4. 트래블소나(Travel-Sona) 모델 제안

소비 성향을 설명하는 네 가지 요인(E/I, S/N, T/F, J/P)은 서로 독립적인 특성을 갖고 있지만, 실제 카드 소비 패턴에서는 복합적으로 작용하는 경우가 많다. 본 연구에서는 이 네 가지 요인의 조합을 바탕으로 총 16가지 소비자 유형을 정리했고, 이 결과를 ‘트래블소나(Travel-Sona)’라는 이름으로 구조화했다.

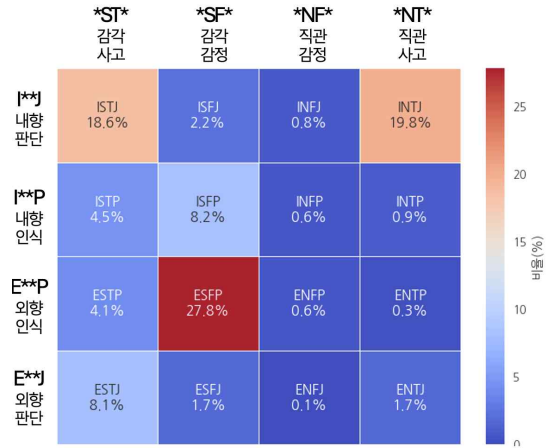


그림 3 소비데이터 분석 결과 트래블소나 모델

트래블소나는 각 소비자의 카드 사용 시간대, 업종 선호, 결제 채널, 여행 방식 등을 종합해서 해당 소비자의 여행 소비 행동이 어떤 유형에 가까운지를 파악할 수 있게 설계하였다. 여기에는 ‘계획형 소비자인가’, ‘즉흥적인가’, ‘가족형?’ 개인형?, ‘신중하게 구매하는가?’, ‘즉흥적인 구매하는가’ 등 소비 이면의 행동 심리를 파악하여 여행 스타일을 추천하고자 하는 목적을 가지고 있다.

예를 들어, E-N-T-P 성향의 소비자는 일반적으로 주말이나 저녁 시간에 온라인을 중심으로 활발히 결제하며, 일정한 규칙보다는 자유로운 의사결정 하에 소비하는 경향을 보였다.

III. 결론

본 연구는 카드 소비 데이터를 바탕으로 개인의 소비 성향을 파악하고, 이를 바탕으로 맞춤형 여행을 추천하기 위한 방법을 제시한다. 이 MBTI의 네 가지 성향 축에서 착안해서 총 16개의 유형을 도출하고, 이를 ‘트래블소나’ 모델로 정리하였다. 다만, 이번 분석에 사용된 소비데이터는 특정 카드사 자료에 국한되어있으며, 데이터의 소비 기간도 길지 않아서 전체적인 소비 성향을 온전히 반영하기에는 제약이 있다. 또한, 일부 거래에는 가족카드나 간편결제 등의 해석이 힘든 내역도 존재한다. 보다 정밀한 모델로 확장하기 위해서는 통합 카드 소비데이터로 확대할 필요가 있다.

ACKNOWLEDGMENT

본 논문은 한국과학기술정보연구원 자체사업(J24JR007-24)의 지원에 의해 이루어진 것임.

참 고 문 헌

- [1] Khatri, J., Marín-Morales, J., Moghaddasi, G. J., Giglioli I. A. C., Alcañiz, M. “Recognizing Personality Traits Using Consumer Behavior Patterns in a Virtual Retail Store”, vol. 13-2022, 2022.
- [2] 석관호 외,“대학생 소비자의 MBTI 유형과 소비자행동의 연관성에 관한 연구: 구매행동 특성, 소셜네트워크 서비스(SNS) 이용동기, 소비자 라이프스타일을 중심으로”, 소비문화연구, vol.26, no.4, pp.19 - 51, 2023.