

머신 러닝을 활용한 의류제품의 판매량 예측 모델 - 아우터웨어 품목을 중심으로 -

채진미[†] · 김은희¹⁾

한성대학교 글로벌패션산업학부

¹⁾한국 오라클

Sales Forecasting Model for Apparel Products Using Machine Learning Technique - A Case Study on Forecasting Outerwear Items -

Jin Mie Chae[†] and Eun Hie Kim¹⁾

School of Global Fashion Industry, Hansung University; Seoul, Korea

¹⁾Oracle Korea

Abstract: Sales forecasting is crucial for many retail operations. For apparel retailers, accurate sales forecast for the next season is critical to properly manage inventory and plan their supply chains. The challenge in this increases because apparel products are always new for the next season, have numerous variations, short life cycles, long lead times, and seasonal trends. In this study, a sales forecasting model is proposed for apparel products using machine learning techniques. The sales data pertaining to outerwear items for four years were collected from a Korean sports brand and filtered with outliers. Subsequently, the data were standardized by removing the effects of exogenous variables. The sales patterns of outerwear items were clustered by applying K-means clustering, and outerwear attributes associated with the specific sales-pattern type were determined by using a decision tree classifier. Six types of sales pattern clusters were derived and classified using a hybrid model of clustering and decision tree algorithm, and finally, the relationship between outerwear attributes and sales patterns was revealed. Each sales pattern can be used to predict stock-keeping-unit-level sales based on item attributes.

Key words: sales forecasting (판매예측), k-means clustering (k-평균 군집분석), decision tree classifier (의사결정나무), sales pattern (판매패턴), outerwear item attributes (아우터웨어 속성)

1. 서 론

적절한 생산량 산정과 재고관리는 패션산업에서 기본이 되는 영역이다. 판매량 예측이 정확하지 않아 생기는 공급과잉 이슈는 재고량 증가로 연결돼 기업의 수익성에 부정적인 영향을 미칠 것이며 반대로, 품절 상황이 발생하면 판매기회 상실이나 고객서비스 저하로 인한 고객의 상표이탈이 발생할 수 있으므로 미래 수요에 대한 판매량의 정확한 예측은 매우 중요하다.

그러나 패션산업은 다른 산업에 비해 판매량 예측이 어렵다. 다음 시즌의 상품에 대한 판매량 예측은 원자재 조달 계획 등의 이유로 6개월~1년 전에 진행되므로 예측 기간이 길고 다음 시즌의 거의 모든 패션상품이 새로운 상품으로 출시되므로 예측모델을 생성하는데 참조할 과거 데이터를 어떤 것을 쓸지 어

려운 상황에 처한다(Vashishtha et al., 2020). 판매량 예측을 더욱 어렵게 하는 것은 품목이 아니라 가장 최소단위인 상품 별로 정확한 판매량 예측이 필요한 점이다. 패션상품의 특성상 다른 상품에 비해 긴 생산 리드타임, 짧은 판매기간을 가지며 'See now buy now' 구매행동에 따라 판매 시점이 되면 소비자 취향과 유행이 달라져 있어 시차에 따른 변동성이 크다. 또한 날씨, 마케팅 전략, 타사의 경쟁상품의 가격뿐만 아니라 거시적 환경 등의 외부요인에 의해 영향을 받는다(Lee et al., 1997). 최근에는 소비자들의 욕구가 다양해짐에 따라 제품의 수명주기가 더욱 짧아지고 있으므로 시장 수요에 대한 정확한 예측이 매우 중요한 문제로 부각되고 있다.

대부분 패션기업에서 신제품에 대한 판매량 예측을 전문가들의 경험이나 과거의 판매 추이를 토대로 기초 통계 분석에 의거해 상품기획자가 엑셀 수작업으로 진행하고 있는데 예측의 정확도가 낮으며 개인의 역량에 좌우되므로 판매량 예측 자동화 시스템 기반으로 운영하는 기업에 비해 경쟁력이 떨어지며 빠른 시장에 대처하기 어렵다. 최근에는 한국의 패션기업들도 데이터 분석 플랫폼을 도입해 판매량 예측 모델을 개발하는 경우가 늘어나면서 모델의 정확도를 높이기 위해 기업의 R&D는 늘어나고 있지만 의류학계와 산학협동 연구는 적은 것이 사실이다.

[†]Corresponding author; Jin Mie Chae

Tel. +82-2-760-8037

E-mail: chaemie61@hanmail.net

© 2021 (by) Fashion and Textile Research Journal (FTRJ). This is an open access journal. Articles are distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

패션상품을 대상으로 수요예측을 분석한 연구를 살펴보면, Lee(2008)는 특정 패션기업의 5년간 판매량 데이터를 대상으로 다섯 종류의 시계열 모형을 이용하여 분석 기법을 비교, 평가하였고, Kim(2009)은 특정 백화점의 3년간 판매량 데이터를 대상으로 매출에 영향을 미치는 요인을 선정하고 회귀분석을 통해 수요예측 모델을 제시하였다. Lee(2012)는 패션시장규모를 추정한 데이터로 지수평활법과 다변량 회귀분석법을 이용하여 남성성장, 여성성장, 캐주얼복, 스포츠복, 외투, 내의, 아동복의 세분시장 별 수요예측 모형을 제안하였고, Lee et al.(2014)은 남성 브랜드의 2년간의 판매량 데이터를 이용하여 의류제품의 수명주기를 예측하는 분석을 시행하였다. 이상의 연구들은 전체 복종 별 판매 데이터에 의거하여 판매량 예측을 시행하였으므로 이 분석 결과를 각 스타일별 판매량 예측의 근거로 적용하기엔 예측력의 한계가 크다. 의류상품의 경우 각 복종 내에서도 많은 품목(item)이 존재하며 각 품목별로 다양한 스타일이 기획, 출시되므로 매 시즌 출시되는 각 스타일의 판매량에 위의 연구들에서 제시한 전체 복종의 예측량을 적용하기엔 무리가 있다. 반면, 신규 상품에 대해 참조할 과거의 판매패턴을 선택할 때 과거 한 개 상품의 판매 패턴을 가져와 적용해 예측을 하는 것도 불합리하므로 신규 상품과 유사한 판매패턴으로 팔릴 것으로 예상되는 유사 상품군을 선택해야 하는데 무슨 기준으로 유사 상품군을 선택할 것인지 어려움이 있다. 이에 대한 해결방법으로 Thomassey et al.(2003)은 품목(item)별 과거의 데이터를 분석하여 각 상품별 판매패턴 예측 시 참조할 수 있는 서브품목(sub-item)을 도출할 것을 제안하였다. 단, 도출한 판매패턴이 차기 시즌에도 유사한 양상을 보일 것이라는 전제를 한다는 한계가 있으므로 이에 대한 해결책으로 본 논문에서는 판매패턴의 표준화를 위해 변동성에 영향을 줄 것으로 예측되는 영향요인을 제거하는 보정작업을 거쳐 최대한 예측 오차를 줄이고자 하였다.

그러므로 본 연구에서는 국내 스포츠의류 제조업체인 K브랜드의 FW 시즌에 가장 대표적인 품목인 아우터웨어(outer wear)를 선정하여 판매량 예측모델을 생성하는 것을 목적으로 하였다. 과거 4년 동안의 판매량 데이터를 근거로 신상품 예측에 활용할 수 있는 표준 판매패턴 모델을 생성하기 위해 데이터 보정작업으로 변동성 영향 요인을 제거한 후 판매패턴의 유형을 도출하고, 각 판매패턴 유형과 연관된 아우터웨어 속성변수를 밝히고자 하였다. 이를 통해 산업현장에서 상품기획자가 상품의 판매량 예측 시 품목 수준에서 여러 개의 참조할 판매패턴이 있고 각 판매패턴과 연관된 상품속성이 무엇인지 알고 있다면 차기 시즌의 같은 상품속성을 갖는 신상품에 대해 보다 정확하게 판매추이를 예측할 수 있는 근거가 될 것이다.

2. 이론적 배경

2.1. 패션산업에서의 판매량 예측을 위한 분석방법

일반적으로 판매량 예측은 크게 정성적 방법(qualitative

method)과 정량적 방법(quantitative method)의 두 가지 방법으로 이루어질 수 있다. 정성적 방법은 주로 분석자의 과거의 경험이나 사전지식(domain knowledge)과 같은 주관적인 의견이나 정보에 기반하여 예측하는 것이고, 정량적 방법은 판매량 데이터와 같은 정량적인 자료를 기반으로 분석방법을 이용하여 예측하는 것이다. Lee et al.(2014)은 정성적 방법을 전문가 기반 예측방법과 시장 기반 예측방법으로 분류했는데 전문가 기반 예측방법은 전문가들의 직관이나 의견에 기반을 두어 예측하는 것이다(Koo & Min, 2013). 시장 기반 예측방법은 시장조사법(market research)과 수명주기 유추법(life cycle analogy)으로 분류되는데, 수명주기 유추법이란 제품의 과거의 수명주기 정보를 이용해서 미래 수요를 추정하는 방법이다(Kim et al., 2011).

Liu et al.(2013)은 판매량 예측은 패션 산업에서 매우 중요한 문제인 만큼 지금까지 예측방법에 대한 많은 연구가 이루어져 왔는데 각각의 방법이 한계점이 있다고 지적하면서 정량적 분석 방법을 통계적 방법에 기반한 예측법, AI(Artificial Intelligence) 기반 예측법, 다양한 유형의 하이브리드(hybrid) 예측법으로 분류하여 설명하였다. 전통적으로 선형회귀분석(linear regression), 이동평균법(moving average), 지수평활법(exponential smoothing), 베이즈 분석(bayesian analysis) 등과 같은 다양한 통계기법과 ARIMA(Auto Regressive Integrated Moving Average)와 SARIMA(Seasonal ARIMA)와 같은 시계열분석 통계기법이 이용되어 왔다(Mostard et al., 2011; Thomassey et al., 2003; Yelland & Dong, 2014). 통계적 분석 방법은 AI 분석법에 비해 비교적 간단하고 컴퓨팅 시간 면에서 빠르게 분석 결과를 확인할 수 있으나, AI 기반 분석법보다 예측율이 낮고 특히 패션 판매량은 계절적 요인이나 트렌드 요인 등 많은 외부 변수들의 영향을 받아 불규칙성을 보이므로 만족할 만한 예측성고를 기대하기 어렵다(Liu et al., 2013).

컴퓨터 기술의 진화로 AI기반 분석법은 예측력을 향상시켰고 인공지능 분석법 중에서 ANN(Artificial Neural Network), ENN(Evolutionary Neural Network). 퍼지로지 모델(fuzzy logic mode)이 패션 판매예측에 많이 이용되어 왔다(Au et al., 2008; Frank et al., 2003; Hui et al., 2005; Sztandera et al., 2004). 이와 같은 모델은 전통적인 ARIMA, SARIMA 모델보다 예측력은 우수하지만 분석 플랫폼과 컴퓨팅 파워가 준비되어야 하고 좀 더 시간이 소요되므로 비교적 간단한 머신러닝 기법인 ELM(Extreme Learning Machine)이나 EELM(Extended ELM) 모델이 개발되었다(Sun et al., 2008; Yu et al., 2012). 머신러닝과 인공지능 기반의 예측법도 여전히 완벽한 모델이라 할 수 없으므로 패션 판매량 예측을 향상시키기 위하여 다양한 유형의 하이브리드 모델이 시도되고 있는 중이다(Liu et al., 2013).

하이브리드 모델이란 다양한 분석기법을 함께 사용하여 예측모델을 개발하는 것인데, 통계분석, 머신러닝, AI 인공지능망의 기법에서 한 가지 분석 기법만 사용하는 것보다 효율성이 높은 것으로 나타났다. 그러므로 여러 학자들이 각각의 분석법의 강점을 조합하여 패션 판매량 예측을 위한 새로운 모델을

제안하였는데, Vroman et al.(1998)은 퍼지로지적 근거한 하이브리드 패션 예측모델 개발의 선구자로 지수평활법 중의 하나인 Holt-윈터(Holt-Winter)모델에 퍼지로직을 조합한 분석법을 개발하였다. Yesil et al.(2012)도 순수한 통계법과 퍼지로직 모델을 조합한 하이브리드 모델을 개발하여 패스트 패션 예측에 적용하였고 예측력이 향상되었다고 설명하였다. 또한 Vroman et al.(2001)은 NN(Neural Network, 인공신경망)을 활용한 하이브리드 모델을 개발하여 단기의 불연속적인 시계열 데이터를 분석하는데 적합하다고 제안하였고, Thomassey and Happiette (2007)는 신제품을 예측하기 위해 클러스터링(clustering)과 분류기법(classification)을 조합한 하이브리드 모델을 제안하였다. Aksoy et al.(2012)은 퍼지로지 분석과 신경망 분석의 장점을 조합하여 하이브리드 모델을 개발했는데 신경망이 다양한 은닉층(hidden layer)으로 인해 설명이 어려운 것에 반해 설명력을 높인 모델이라고 설명하였으며, Choi et al.(2012)은 ANN과 GM(Grey Model)을 조합한 하이브리드 모델을 개발하여 컬러 예측을 진행하였다. 이외에도 Thomassey and Fiordaliso(2006)는 군집화(clustering)와 의사결정나무 분류법(decision tree classifier)을 조합한 하이브리드 모델을 개발하였고, Ni and Fan(2011)은 자기회귀(autoregression)와 의사결정나무(decision tree)를 조합한 분석법을 개발하여 패션상품 판매예측에 매우 적합하다고 제안하였다.

이와 같이 많은 연구에서 패션상품의 판매예측력을 높이기 위해 다양한 분석기법들을 제안해 왔는데, 분석 시에는 데이터의 양이나 변수 특성(범주형, 숫자형), 향후 얼마 후의 예측인자의 기간을 고려해야 하고(Makridakis & Wheelwright, 1978), 최근에는 이미지, 텍스트 등 데이터가 분석에 중요해지면서 데이터 종류(구조형 데이터, 빅데이터), 예측에 필요한 분석 플랫폼의 인프라 상황, 전문가 지식, 분석 기법 별 예측 정확도 등 여러 요인을 고려하여 가장 적합한 분석법을 선택해야 할 것이다.

2.2. 군집화(clustering)와 의사결정나무(decision tree)에 기반한 하이브리드 모델

패션산업에서 판매예측은 매우 복잡한 문제이다. 매 시즌마다 방대한 수의 스타일이 생산되며, 새로운 상품으로 교체되므로 신제품 예측 시 참조할 과거의 동일한 판매 데이터가 존재하지 않는다. 일반적으로 상품의 수명주기도 짧으며 다양한 변수들, 즉 컬러와 가격 같은 상품 특성, 점포 수와 같은 유통상황, 유행, 경제상황, SNS, 날씨, 휴일 효과와 같은 외부 요인뿐만 아니라 소비자들의 구매심리 등과 같은 때로는 측정 가능하지 않은 변수들까지도 고려되어야 한다(Thomassey & Fiordaliso, 2006). 예측모델을 개발하기 위해서는 과거 몇 년 동안의 누적된 데이터가 필요한데 판매패턴이 주기적으로 반복되는 특정 산업영역에서는 예측모델의 결과가 좋지만, 과거의 누적된 동일한 판매 데이터가 존재하지 않는 패션기업의 신제품 판매예측에 대해서는 한계가 있다고 할 수 있다.

그러나 Saaksvuori and Immonen(2005)은 매 시즌 새로 기

획되는 신상품의 경우 과거의 동일한 판매 데이터는 없으나 비슷한 스타일의 제품이 이전 시즌에서 판매되었을 것이고, 신상품의 판매예측을 위해 과거에 판매되었던 비슷한 스타일의 판매 데이터를 사용할 수 있을 것이라고 하였다. 그러나 실제 현업에서 판매 데이터를 분석해 보면, 시각적 특징이 유사한 스타일이 과거의 스타일과 동일한 판매패턴을 보이지 않는 경우가 많다. 따라서 유사 상품을 정의할 때 시각적으로 유사하다고 판단되는 스타일보다는 유사한 판매패턴으로 분류하는데 연관된 상품 속성을 밝혀낸 후, 예측 시에는 연관된 상품의 속성을 사용할 수 있을 것이다.

이러한 개념을 전제로 하여 본 연구에서는 Thomassey and Fiordaliso(2006)가 제시한 군집화와 의사결정나무의 하이브리드 모델을 적용하여 판매예측 모델을 제안하고자 한다. 본 연구 모델의 장점은 군집화를 통해 과거의 다량의 판매데이터를 유사한 형태를 나타내는 집단들로 분류하여 판매패턴의 유형을 도출할 수 있다는 점이다. 의사결정나무는 명목변수나 연속변수 등에 구애받지 않고 데이터를 사용할 수 있으므로 의류 속성변수 처리에 적합하다. 분석을 완료한 후에는 예측값(Y)에 대한 영향변수의 설명이 필요한 경우가 많은데, 판매패턴 데이터를 군집화한 후에 각각의 사례(Y)를 변수(X)를 활용해 그것이 속한 군집에 할당(예측, Predicted Y)하는 기법인 의사결정나무를 적용해 각 군집 별 변수의 설명력을 분석한다(Witten & Frank, 1999). 동일 집단으로 분류하는 분석법 중에서 k-평균 군집분석은 X변수만 사용해 군집하는데 효과적이며(Thomassey & Fiordaliso, 2006), 의사결정나무는 베이즈망, 신경망 분석과 함께 X변수, Y변수가 모두 주어진 상태에서 동일 집단으로 분류하는데 가장 적합한 분석법으로 이용되어 왔다(Tsujino & Nishida, 1995). 특히 의사결정나무는 사용하기에 간단하고 이해하기 쉽게 분류하므로 매우 보편적인 분석법이며(Lee & Oh, 1996; Tsujino & Nishida, 1995). 변수의 정규성, 등분산성, 선형성이 전제되지 않아도 사용할 수 있다는 편리함이 있다. 따라서 의류상품의 속성에는 가격 외에도 기획군, 스타일, 재질, 혼용용 등 다양한 범주형 데이터가 혼재되어 있는데 이러한 데이터 유형에 구애받지 않고 분석을 수행할 수 있다.

의류상품의 실증적 판매 데이터를 이용하여 판매량 예측모델을 연구한 국내 연구 사례는 극히 드물다. 의사결정나무 기법을 이용한 국내 연구로는 Lee et al.(2014)의 연구가 있는데, 남성 브랜드의 500여개의 제품을 포함한 2년간 데이터를 이용하여 예측모델을 생성하고 새로운 제품의 수명주기 패턴을 예측하였다. 주별 판매량 데이터를 근거로 군집분석 결과 high peak, fast-up, slow-up, never-up의 4종류의 제품수명주기 패턴을 도출하였고, 각 상품의 품목, 가격, 디자인, 색상, 소재, 패턴, 핏(fit)의 7가지 속성정보와 출시된 후 초기 5주간의 판매량 정보를 이용하여 의사결정나무 모델을 구성하였다. 이 연구는 의류상품의 품목을 모두 포함한 데이터이므로 의류상품의 각각의 품목에 대한 판매량 예측에 적용하기에는 한계가 있다는 판단이다.

의류상품의 경우 다른 상품들에 비해 품목이 다양하고 품목 내 다양한 세부 품목이 전개되어 있으며 각 세부 품목에 대해 소비자의 구매행동이 다르다고 판단된다. 한미 FTA의 HS 코드에서도 남성, 여성에 대해서 슈트, 재킷, 셔츠, 블라우스, 티셔츠, 스웨터, 코트, 드레스 등 다양한 코드로 구성되어 있고 각 품목에 대해서도 스타일의 다양한 변형이 존재한다. 예를 들어 세부 품목 중에서 티셔츠와 다운패딩 재킷의 판매 패턴은 날씨의 영향요인, 소비자의 취향, 연령대 등으로 인해 구매 행동이 현저히 다르므로 다른 판매형태를 나타낼 것이다. 그러므로 의류상품의 경우 전체 상품의 판매패턴을 도출해 각 품목에 적용한다는 것은 그 실효성이 떨어질 것이므로 본 연구에서는 품목별로 판매패턴을 도출하고자 하였으며, 국내 K스포츠브랜드 FW시즌의 대표 품목인 아우터웨어(outerwear)를 대상으로 분석하였다.

3. 연구 방법

3.1. 연구 문제

K브랜드의 과거 4년 동안 누적된 아우터웨어 품목의 매출데이터를 대상으로 표준 판매패턴 유형을 군집하고, 특정 판매패턴 유형으로 분류하는데 연관된 품목속성을 도출하여 동일 속성을 가지는 신상품의 판매량 예측모델로 사용하고자 하였다. 이를 위한 구체적인 연구문제는 다음과 같으며, 연구모델은 Fig. 1과 같다.

연구문제 1: 아우터웨어 품목의 판매패턴 유형을 도출한다.

연구문제 2: 특정 판매패턴과 연관된 아우터웨어 속성을 밝히고 판매패턴 유형을 생성한다.

3.2. 데이터 수집과 전처리

K브랜드에서 최근 5년 동안 판매한 FW시즌의 아우터웨어 품목을 대상으로 총 436개 상품 코드의 매출데이터를 수집하였으며, 이 중에서 4년간의 판매데이터로 판매예측 모델을 생성하였다. 예측의 단위는 스타일, 컬러 코드 단계의 SKU(Stock Keeping Unit)로 정의했다. SKU는 재고관리의 최소단위로 각

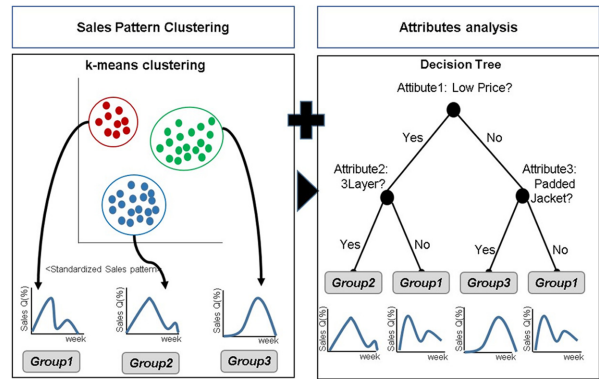


Fig. 1. Research model.

상품을 의미하나 K브랜드에서는 스타일+컬러 단계까지의 상품 코드를 SKU로 사용하고 있으며, 예측도 컬러 단계까지 진행하고 사이즈 단계는 각 사이즈별 판매비율을 사용해 분배하고 있다. 아우터웨어 품목으로는 고어텍스 재킷, 기능성 재킷, 경량 다운재킷, 중량다운재킷 및 그 외 기타 캐주얼재킷 등을 포함하고 있고, 분석 단위는 실제로 FW시즌의 제품이 판매되었던 기간(9월1주~3월4주, 28주)의 일단위 판매량(Q)을 사용하였다. 분석을 위해 SAS의 EG(Enterprise Guide)를 사용해 데이터를 추출하였으며 분석 플랫폼인 SAS의 E-miner를 사용해 회귀분석을 이용하여 데이터 표준화를 시행하였다. 연도별 판매추이는 Fig. 2와 같으며, 이에 대한 판매량과 할인율은 Table 1과 같다.

연도별로 판매 패턴을 살펴보면 판매가 증가하면서 10월 중순~11월 중순에 피크를 이루며 시즌 말기에는 할인에 의해 판매량이 증가했다. 5년 동안 판매량을 살펴볼 때 초기 2년은 시즌 말기에 10% 이하의 할인을 했으나 최근 3년 동안은 할인을 자주 진행했고 평균 20% 이상의 할인을 시행했음에도 불구하고 급격한 판매량의 증가가 이뤄지지 않는 양상이다. 판매량을 살펴보면, 초기 3년간 감소 추세였다가 추세가 반전해 판매량이 증가했으며 최근에는 기획 상품스타일 전개 코드 수가 감소했지만 판매량이 증가한 것으로 비추어 호조판매 상품에 선택

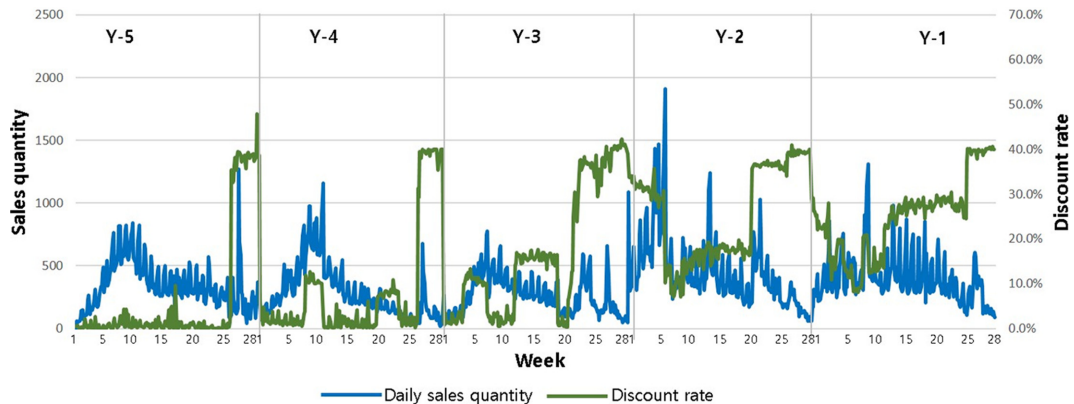


Fig. 2. Sales pattern of outerwear.

Table 1. Sales data of outerwear

Year	Y-5	Y-4	Y-3	Y-2	Y-1
Cumulative sales quantity	70,322	57,910	53,180	91,113	79,905
Average discount rate	5.3%	7.0%	16.1%	25.6%	24.6%
Number of SKU	101	99	85	95	118

Cumulative sales quantity = total number of outerwear sales quantity
 Number of SKU = total number of style quantity

과 집중을 한 상품 구색 기획이라고 판단할 수 있다. 신상품 예측 시 연도 간 변동성을 줄이고 이상치(outlier)에 영향을 적게 받는 예측 모델을 만들기 위해 데이터를 제거하고 표준화하는 작업을 시행하였다. 먼저 상품 코드에서 상설매장 판매, 직원판매, 단체판매, 반품된 상품의 데이터를 제외했다. 둘째로 판매패턴의 정상 범위에서 벗어나는 이상치(특이값, outlier) 들을 제외했다. 총판매량이 소량인 상품(300개 미만)은 판매패턴의 파동이 심하기 때문에, 조기완판/재고부족 상품은 판매패턴이 초기에 종료되므로, 입고지연되거나 스팟생산된 상품은 판매패턴이 중기 이후에 시작되므로 모두 제외한 후, 총 104개 SKU를 사용하였다. K브랜드가 대형 브랜드이지만 다품종 소량생산이 많고, 아우터웨어 품목 중에서 1000장 이상 생산하는 주력 SKU의 판매량예측 모델을 생성하고자 하는 것이므로 다수의 SKU를 제외하는 작업을 거쳤다.

다음으로 판매에 변동성의 영향을 주는 외부변수에 의한 판매량 변동분을 제거하는 데이터 표준화(standardization)를 시행하였다. 첫 번째 단계로 판매량에 영향을 준 외부변수를 확인하였다. 의류제품의 판매량에 영향을 주는 요인으로는 날씨요인(Arunraj & Ahrens, 2016; Bahng & Kincade, 2012; Jang & Lee, 2002), 휴일효과(Nam, 2006; Hwangbo et al., 2017), 가격할인(Jang & Lim, 2003; Kim & Hwangbo, 2017; Sung, 2006) 등의 요인이 밝혀져 왔다. 그러므로 본 연구에서도 이들 요인들의 영향력을 확인하고자 기온효과, 휴일효과, 가격할인을 독립변수로 투입하고 판매량에 미치는 영향을 회귀분석을 통해 분석한 결과, 날씨요인은 유의한 p 값(p value)이 나오지 않았으며 휴일효과와 가격할인이 유의한 변수로 확인되었다. 두 번째 단계로 휴일효과와 가격할인으로 인한 판매량의 변동분을 제거하는 표준화 작업을 시행하였다.

가격할인과 휴일효과의 일별 판매량(Q)에 대한 영향력은 시즌과 가격의 6가지 조건에 따라 6가지 회귀계수(coefficient)를 계산하였다. 시즌은 초기(9월 1주~11월 2주), 중기(11월 3주~1월 2주), 후기(1월 3주~3월 3주)의 3가지로 구분하고 가격은 고가(40만원 이상), 저가(40만원 미만)로 구분하여 6가지 조건에 대해서 각각 회귀분석 후 회귀계수를 도출해 영향력(sensitivity)을 계산하였다. 회귀분석에서 X변수는 요일, 휴일, 주, 연도, 할인을 입력하고 Y변수는 일별 판매율(%)을 입력하였다. 가격할인 민감도(α)는 1% 할인이 증가할 때 증가하는 일별 판매율(%)로 정의하였으며, 휴일효과 민감도(β)는 휴일이 있을 때 증가하는 일별 판매율(%)로 정의하였다. 마지막으로 보정한 일별

판매량을 계산하였는데, 현재 일별 판매량에서 위에서 제시한 할인, 휴일 효과에 의한 일별 판매량을 빼서 보정한 일별 판매량을 계산하였고, 이를 주별 판매량으로 합하여 주별 데이터를 사용하였다. 외부효과를 제거하기 위해 사용된 보정식은 Eq. 1과 같다.

$$\hat{y}_t = \frac{y_t}{1 + \alpha \times DCrate + \beta \times Holiday} \quad \text{Eq. (1)}$$

\hat{y}_t : Calculated daily sales quantity at the point of time t

y_t : Daily sales quantity at the point of time t

α : Coefficient of discount sensitivity

β : Coefficient of weekend sensitivity

3.3. 분석 방법

본 연구에서는 품목 내에서 유사 판매패턴의 유형을 도출하기 위해 군집화 알고리즘을 사용하였으며, 판매패턴의 유형 분류 시 예측률이 초점인 알고리즘(신경망, SVM, Naive Bayes, 랜덤 포레스트)보다는 분류 기준을 제시해주는 알고리즘인 의사결정나무를 선택하였다. 첫 번째로 아우터웨어 품목의 판매패턴 유형은 각 SKU의 판매패턴(X)을 거리의 가까움 정도(유사성)에 따라 군집화하는 K-평균 군집분석을 수행해 분류하였다. 최적의 K값(군집수)은 SKU의 주별 판매량 데이터를 입력하여 데이터 사이 거리의 합이 최소가 되는 값으로 결정하였다. 클러스터링을 반복 시행한 후에 최적의 군집 수인 K값은 pseudo t^2 값, CCC(Cubic Clustering Criteria), ABC(Aligned Box Criterion) 값을 비교해서 결정하였으며, pseudo t^2 값이 최소화되도록 선정하였다. 클러스터링은 X변수만 사용해 거리가 가깝게 군집하는 머신러닝의 비지도학습(unsupervised learning)이다(Bhavsar et al., 2017). 통계적 분석방법이 표본 샘플링과 가설검증에 초점을 맞추는데 비해 머신 러닝은 전체 데이터를 대상으로 알고리즘의 매개변수를 학습을 통해 발견해 내는 것을 특징으로 한다.

두 번째로 아우터웨어의 판매패턴 유형들과 연관된 품목속성을 찾아내기 위하여 앞서 도출된 군집을 목표값(Y)으로 정의한 후, 속성(X)과의 연관성을 밝혀 유의속성(X_p)과 속성의 분할기준($X_p > c$)을 찾아내는 의사결정나무를 생성하였다. 의사결정나무는 Y변수가 주어졌을 때, 유사한 군집으로 분류되도록 X변수의 조합을 찾아내는 분석법이며(Müller & Guido, 2016), 머신러닝에서는 알고 있는 결과(Y, label)를 가지고 학습한 후에 이를 기반으로 유사한 집단으로 분류(classify)하는 지도학습(supervised learning)에 속한다(Bhavsar et al., 2017). Hastie et al.(2001)은 의사결정나무는 다른 분석법에 비해 범주형과 연속형 데이터를 모두 사용가능하다고 하였고, Curram and Mingers (1994)는 분석된 결과를 이해하고 해석하기 쉬운 분석법이라고 하였다. 본 연구에서 투입된 X변수는 아우터웨어 품목의 가격(연속형 변수) 외에도 레이어종류, 스타일, 혼용율, 기획군 등 다양한 범주형 변수가 혼합되어 있고, 분류된 판매패턴 유형과

연관이 있는 X변수를 설명하기 위해서 의사결정나무가 가장 적합하다고 판단되었다.

4. 연구 결과

4.1. 아우터웨어의 판매패턴 유형 도출

아우터웨어 품목의 SKU별 판매데이터를 K-평균 군집분석을 한 결과 반복적인 클러스터링을 통해 7개의 유형을 도출하였으며(Table 2), 가능한 범용적인 판매패턴을 추출하는데 목적이 있으므로 군집 4를 이상치로 보고 제외하였다. 이상치를 탐지하기 위해서 사분위수 범위를 선정하는 방법을 쓰는데(Joo &

Cho, 2016), 군집 4는 3사분위수에 포함되므로 이상치로 간주하였다.

이상치로 판단한 군집 4를 제외한 6개의 판매패턴 유형을 도출하였고 총 100개의 SKU가 포함되었다. 각 SKU별 판매패턴과 그의 평균을 시각화한 표준판매패턴은 Fig. 3과 같다.

표준판매패턴 유형에 대해서 설명하면, 유형 5는 가장 많은 SKU(34개, 32.7%)를 포함한 판매패턴으로 초기시즌인 9월 4주부터 급격히 매출이 상승하여 10월 3주에 정점을 이룬 후 빠르게 하강하는 삼각형(triangle) 형태의 그래프이며, 유형 6과 비슷한 형태를 보인다. 유형 3과 유형 7은 다음으로 많은 SKU를 포함한 판매패턴인데, 형태는 초기시즌인 9월 4주부터 조금씩 매출이 상승하다가 11월 2주에 매출이 정점을 이루는 삼각형 그래프이다. 두 가지 패턴의 차이를 살펴보면 유형 5와 6의 상품들이 먼저 판매량이 급상승하고 정점도 빠른 형태를 보이는데, 이를 통해 유형 3과 7에 비해 상품의 두께감이 얇고 계절성이 이른 것으로 추측할 수 있다. 이러한 상품속성은 의사결정나무를 통해 확인해 볼 수 있다. 판매패턴 유형 1은 중기시즌 11월 2주부터 판매가 상승하여 12월 3주에 매출이 정점을 이루는 것으로 보아 상대적으로 두께감이 있는 겨울용 상품으로 추론할 수 있으며, 유형 2는 초기부터 후기시즌인 1월 4주까지 꾸준히 판매가 발생하는 플라토(plateau) 형태를 보인다.

Table 2. Number of SKU for each cluster

Cluster No.	SKU #	Ratio(%)
Cluster 1	7	6.7%
Cluster 2	13	12.5%
Cluster 3	20	19.2%
Cluster 4	4	3.8%
Cluster 5	34	32.7%
Cluster 6	7	6.7%
Cluster 7	19	18.4%
Sub total (of 7 Cluster)	104	100%

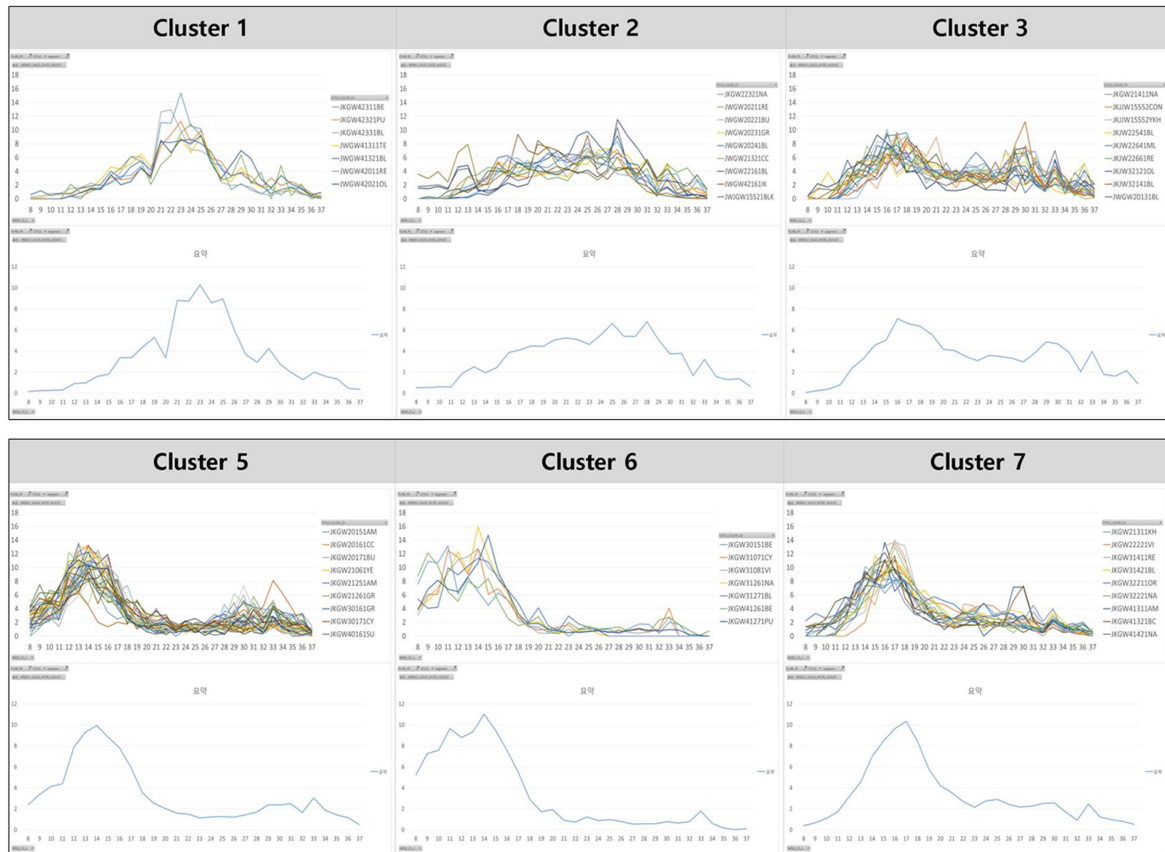


Fig. 3. Sales patterns and average sales pattern for each cluster.

6개 유형을 볼 때 판매패턴의 형태는 매출이 장시간 크게 발생하는 종형(bell)은 없으며 대부분의 형태가 정점까지 매출이 상승하다가 정점 이후에는 하락하는 삼각형(triangle) 형태이며, 시즌 후반기에 매출이 한 번 더 약간 상승하는 형태를 볼 수 있다.

이를 통해 K브랜드가 판매하고 있는 FW 아우터웨어 품목은 9월 4주부터 매출이 증가하면서 가을에 정점을 이루는 상품(유형 5, 유형 6), 초겨울에 정점을 이루는 상품(유형 3, 유형 7), 성동기 12월 4주~1월 2주에 정점을 이루는 상품(유형 1, 유형 2)과 비교적 오랜 기간 동안 꾸준히 판매되는 상품(유형 2, 유형 3)으로 다양한 판매패턴이 도출된 것을 알 수 있다. 이는 K브랜드의 아우터웨어 품목의 특성상 고어재킷부터 초경량다운, 중량다운 재킷, 성동패딩 등 다양한 스타일을 기획하여 판매하고 있으므로, 소비자들이 날씨의 변화에 따라 집중적으로 구매하는 품목이 달라짐을 보여주고 있으며, 이에 따라 품목 별로 판매패턴이 다양하게 도출된 것으로 추론된다.

4.2. 판매패턴 유형에 연관된 아우터웨어 속성

도출한 6개의 판매패턴 유형을 각 SKU별 데이터에 칼럼을 생성하고 입력한 후에 의사결정나무 모델을 생성했다. 의사결정나무에서는 판매패턴 유형의 칼럼을 Y값(목표치, 예측치)으로 설정했으며 유사한 그룹으로 분류할 때 사용하는 아우터웨어 속성을 찾아내며 유사그룹으로 분류된 SKU의 평균 표준판매패턴이 앞서 클러스터링한 평균 표준판매패턴과 비슷하게 나올 수 있도록 하는 것을 목표로 했다. 아우터웨어 속성은 K브랜드에서 아우터웨어 상품 기획 시 고려하는 속성기준들을 모두 투입하여 이 중에서 과연 특정 판매패턴 유형으로 분류하는데 고려할 수 있는 가장 적합한 속성기준을 찾고자 하였다. 착용용도(use), 기장(length), 레이어(layer), 내피(liner), 스타일(style), 혼용율(fabric composition), 기획그룹(plan group)의 속성 기준을 투입하였으며, 최적의 의사결정나무를 생성하기 위해서 노드의 가지치기, 가지의 깊이 조정 등의 매개변수를 변

경해 가면서 의사결정나무를 반복해서 생성하였다. 마지막 노드로 분류된 집단이 동질하고(SKU별로 입력된 클러스터 유형이 같고), 오분류율이 가장 낮은 의사결정나무를 최종적으로 선택하였다. 분석결과 아우터웨어 품목의 판매패턴 유형을 분류하는데 관련된 속성은 스타일과 기획그룹의 2개 속성으로 밝혀졌다(Table 3).

생성된 의사결정나무 Fig. 4를 살펴보면 PCluser(Prediction cluster, 예측값)는 동질한 그룹으로 분류했을 때 예측된 클러스터를 의미하며 노드 1번과 노드 5번의 SKU들은 3번 클러스터 유형으로 예측되었음을 의미한다. X변수를 사용해서 분할하고 동질하게, 오분류율이 최소가 되도록 의사결정나무를 생성했으므로 노드 1번에는 대부분이 3번 클러스터로 입력된 SKU(스타일)가 포함되어 있다. 의사결정나무를 반복해서 생성했으며 최종 의사결정나무 결과에서 볼 수 있듯이 예측 클러스터의 1, 2, 3, 5, 7번 유형은 포함되었고 사전 정의된 클러스터 유형 6번은 속성과 연관성을 찾지 못하고 탈락되었다.

의사결정나무에서 도출한 각 판매패턴과 아우터웨어 속성의 분할기준은 Table 4와 같으며 아우터웨어 전체 스타일을 속성과 연관성을 설명하면서 6개의 유형으로 분류되었다. 또, 분류된 집단의 평균 판매패턴을 그려보면 앞서 클러스터링 유형의 평균판매 패턴과 거의 유사하게 그래프가 도출되었다.

Table 4에서 제시한 판매패턴과 연관된 아우터웨어 속성을 살펴보면, 두께감과 계절성의 상품속성과 판매패턴의 정점의 시기가 연관된 것으로 생각할 수 있다. 예를 들어, 유형 5는 가장 얇은 ZIZ(zip in zip) 고어재킷과 경량패딩을 포함하며 판매패턴의 형태는 10월 3주에 정점을 이루는 그래프이다. 유형 3은 중간 두께감의 중량패딩과 폴리에스터내피 재킷을 포함하며 유형 7은 다운내피 재킷을 포함하는데 두 가지 유형 모두 11월 2주에 정점을 이루는 그래프이다. 유형 1은 가장 두껍고 방한용인 성동패딩을 포함하고 있으며 형태는 12월 3주에 정점을 이루는 것으로 나타났으므로 얇은 ZIZ 고어재킷과 경량패딩은 먼저 매출이 발생하여 겨울초기에 판매가 정점을 이루고 두꺼

Table 3. Outerwear attributes used in decision tree modeling

Attribute	Attribute value	Relevance
Use	EX, TRK, TOWN	X
Length	Regular, Half, Long	X
Layer	2L layer, 2L padded layer, 3L layer	X
Liner	None, Gore tex zip-in-zip liner, Down padded liner, Polyester padded liner	X
Style	Jacket(winter breaker jacket), ZIZ(zip in zip) gore jacket, Down liner jacket, Polyester liner jacket, Light down padded jacket, Medium down padded jacket, Heavy down padded jacket,	O
Composition	PA 100, PL 100, N/P, N/P, N/P/PU, P/N	X
Plan_group	JW, JK	O
Price	High price, Low price	X

EX; expert line, TRK; trekking, Town; casual, PA; nylon(polyamid), PL; polyester, N/P; nylon/polyester, PU; polyurethane, JW; men's line, JK; women's line, 2L Layer; Gore tex, 2L layer construction=face fabric+membrane, 3L Layer; Gore tex 3L layer construction=face fabric+membrane+knit backing, ZIZ; zip in zip

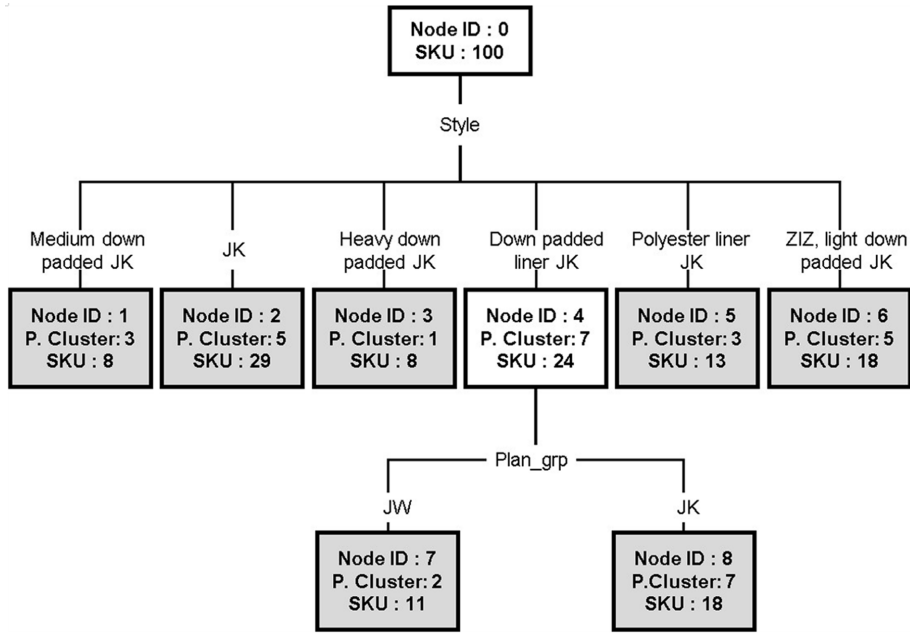


Fig. 4. Decision tree result.

Table 4. Outerwear attributes value related with predicted cluster

Cluster	Style	Splitting attribute 1	Splitting attribute 2	Sales peak
1	Style	Heavy down padded jacket		Dec. 3 rd week
2	Style	Down padded liner jacket	Plan_Grp JW	Plateau
3	Style	Medium down padded jacket Polyester liner jacket		Nov. 2 nd week
5	Style	Jacket ZIZ gore jacket, light down padded jacket		Oct. 3 rd week
7	Style	Down padded liner jacket	Plan_Grp JK	Nov. 2 nd week

운 성동패딩은 겨울 혹한기에 판매가 정점을 이루는 것으로 생각해 볼 수 있다. 유형 2는 중간 두께감의 남성용 다운내피를 포함했으며 11월부터 1월4주까지 지속적으로 꾸준한 판매가 발생하는 판매패턴을 보였다. 이상의 결과를 볼 때, 스포츠브랜드의 아우터웨어의 판매패턴은 두께감과 내피의 형태를 반영하여 분류한 스타일 속성에 따라 판매추이를 예측할 수 있음을 알 수 있다.

5. 결론 및 제언

K브랜드의 아우터웨어 품목을 대상으로 4년 동안의 FW시즌 데이터를 분석하여 표준판매패턴을 유형화하고 연관된 아우터웨어 속성을 분석한 결과는 다음과 같다.

데이터를 준비하는 과정에서 강건한(robust) 예측모델을 생성하기 위해 판매패턴의 변동성을 증가시키는 SKU의 판매패턴은 이상치로 보고 제외하였다. 제외한 SKU에 속한 상품은 소량판매 상품, 조기완관/재고 부족상품, 입고지연/스팟생산 상품이다. 또한, 해마다 다른 판매패턴의 불규칙성과 편차에 영향을

주는 요인을 제거하기 위해 회귀분석을 실시하여 영향요인으로 할인과 휴일효과를 도출했으며, 이들 변수에 의한 판매량 변동분은 보정식을 통해 제거함으로써 데이터를 표준화하였다. 정제된 데이터로 구성된 SKU별 판매패턴을 K-평균 클러스터링 알고리즘을 사용해서 머신러닝을 진행한 결과 6개 유형의 판매패턴 군집이 도출되었으며 이를 의사결정나무 알고리즘을 사용해서 분류한 결과, 최종적으로 5개의 판매패턴 유형과 이와 연관된 아우터웨어 속성으로는 스타일과 기획그룹이 도출되었다.

도출된 판매패턴 유형과 아우터웨어 속성을 연관해서 살펴보면, 두께감과 계절성이라는 요인이 판매패턴의 형태(shape)와 정점(peak) 시기에 연관된 것으로 생각할 수 있다. 두께감이 얇은 아우터웨어일수록 판매가 일찍 시작되어 매출의 정점에 빨리 도달하였으며, 두꺼운 방한용 아우터웨어일수록 판매가 천천히 완만하게 증가하면서 정점의 시기가 성동기(12월 3주)에 이르고 완만하게 판매가 감소하는 양상이었다. 예를 들어, 두께감이 얇은 ZIZ 고어제킷과 경량패딩은 시즌 초기에 판매가 급격히 상승해서 10월 3주에 정점을 찍은 후 판매가 급속히 감소하게 되는데, 특히 ZIZ 고어제킷은 매년 시즌 초에 팔리는

주요 품목이므로 초기에 충분한 물량을 확보할 필요가 있다. 중간 두께감의 중량패딩, 폴리에스터 내피 재킷, 다운내피 재킷은 11월 2주의 정점까지 상승하다가 하강하는 형태를 보인다. 가장 두꺼운 방한용 성동패딩은 12월 3주의 정점까지 꾸준히 판매가 상승하다가 정점 이후 완만히 하강하는 형태를 보여 12월 이후에도 겨울용 아우터웨어로 판매가 지속적으로 발생함을 알 수 있다. 특히 중간 두께감의 남성용 다운내피 재킷은 11월부터 1월 4주까지 지속적으로 판매가 진행되는 플라토(plateau) 형태를 보이고 있어 시즌 내내 안정적으로 판매가 되는 스테이플(staple) 상품으로 분석되었다.

본 연구의 의의는 스포츠브랜드에서 판매한 실증적 데이터를 토대로 클러스터링과 의사결정나무 알고리즘을 합친 하이브리드 모델을 생성하고 표준화된 판매패턴을 생성하기 위해 데이터 정제에 다양한 시도를 한 점에 있다. 학술적으로는 신규 상품의 판매예측을 할 때 기존의 같은 상품의 판매패턴이 없어서 예측이 불가능했던 것을 보완하고자 연구에서 밝힌 판매패턴-속성 연관성을 토대로 속성이 같은 판매패턴을 적용하여 예측이 가능하도록 한 점, 최근의 예측을 향상을 위해 앙상블 모델을 생성하는 것과 같은 맥락으로 하이브리드 모델을 생성한 점에 의의가 있다. 하이브리드 모델의 예측률이 높아지는 점에 착안하여 비지도학습(clustering)과 지도학습(classification)에 속하는 다양한 알고리즘을 하이브리드로 사용해 머신러닝 모델을 생성할 수 있는 가능성을 보여주었다. 실무적으로는 SKU수준에서 신상품의 판매가 시작되면 6주 이후부터 매주 판매패턴을 적용해서 판매량을 예측할 수 있다. 초기 6주차에는 판매량 데이터가 부족하여 예측률이 낮을 수 있으나 판매주차가 많아질수록 판매량 예측이 좀더 정확해지는 결과를 실무에서 얻은 바 있다. 기존에는 품목별로 동일한 판매패턴을 일률적으로 적용하였으므로 좀 더 세밀한 판매예측이 불가능했고, 단순히 시각적으로 스타일이 비슷한 경우의 판매패턴을 적용하였으므로 예측률의 오차가 높은 결과가 나왔다. 본 연구의 하이브리드 모델을 통해 판매패턴의 동질성이 속성 조합에 따라 결정됨이 밝혀졌으므로 신제품의 판매량을 예측 시, 과거에 판매되었던 같은 속성의 판매패턴을 선택하여 판매량을 예측함에 의해 정확도를 높일 수 있는 효과가 있을 것이다. 산업현장에서 판매예측을 할 때 품목별로 다양한 판매패턴이 존재하는데, 판매패턴의 동질성을 설명하는 것이 시각적으로 유사한 스타일이 아니라 속성 수준에서 속성의 조합이라는 것을 밝힌 점에서 의의를 찾고자 한다. 스포츠 브랜드의 아우터웨어라는 품목의 특성 상 두께감을 반영한 스타일 속성이 판매패턴을 결정짓는 속성으로 밝혀졌으나, 다른 품목의 경우, 예를 들어 티셔츠나 팬츠 등의 품목에서 판매패턴을 결정짓는 속성의 조합은 다를 수 있으므로 후속연구로 제안하고자 한다.

본 연구의 한계점으로는 특정 브랜드의 품목에 한정해서 분석을 진행하였고, 데이터 정제과정에서 특정 데이터에 적합한 방법론을 시도했으므로 일반화하기에는 제약이 있다. 클러스터링과 의사결정나무 알고리즘을 반복 시행하면서 매개변수의 변

경을 통해 최적의 정확도를 가지는 모델을 만들려고 노력했지만, 예측율을 더 높일 수 있도록 그 외 다양한 알고리즘을 사용하고 각 알고리즘별로 반복 실험을 진행할 필요가 있다. 또한, 속성 변수를 수집할 때 기능적 아웃도어재킷에 사용하는 보편적 속성을 사용했으나 컬러, 재질감, 기능성 등 보다 다양한 속성 변수를 추가로 시도하여 좀 더 정교한 속성변수를 도출해 내는 노력이 필요할 것으로 판단된다. 추후 본 연구결과 분류된 판매패턴 유형에 따라 테스트 데이터를 투입하여 예측 성과를 검증함이 필요할 것이다. 또한 의류상품에서 속성보다 더 함축적인 정보를 담고 있는 데이터는 상품자체의 이미지라고 할 수 있다. 상품 이미지를 사용해 판매패턴 유형과의 관계를 분석한다면 판매예측 모델의 예측력을 높일 수 있을 것이라고 생각한다. 예를 들어 인공지능의 CNN(Convolutional Neural Network)을 활용한 이미지와 판매패턴의 관계를 연구한 논문이 소개되었는데(Craparotta et al., 2019), 최근에는 한국 패션기업에서도 이미지를 데이터베이스에 연계저장하고 분석 플랫폼으로 통합해 분석할 수 있는 인프라를 속속 도입하고 있으므로 이미지 데이터 분석이 가능할 것으로 보인다. 이외에도 인공지능에서 이미지 인식을 위한 알고리즘도 진보하고 있으므로 향후에는 정제된 데이터의 준비, 알고리즘 실험의 반복을 통해 좀 더 정확한 판매예측 모델을 생성할 수 있으며 발전된 방법론이 나올 수 있을 것으로 기대한다.

감사의 글

본 연구는 한성대학교 교내학술연구비 지원과제임.

References

- Aksoy, A., Ozturk, N., & Sucky, E. (2012). A decision support system for demand forecasting in the clothing industry. *International Journal of Clothing Science and Technology*, 24(4), 221-236. doi:10.1108/09556221211232829
- Arunraj, N. S., & Ahrens, D. (2016). Estimation of non-catastrophic weather impacts for retail industry. *International Journal of Retail & Distribution Management*, 44(7), 731-753. doi:10.1108/IJRDM-07-2015-0101
- Au, K. F., Choi, T. M., & Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*, 114(2), 615-630. doi:10.1016/j.ijpe.2007.06.013
- Bahng, Y., & Kincade, D. H. (2012). The relationship between temperature and sales - Sales data analysis of a retailer of branded women's business wear. *International Journal of Retail & Distribution Management*, 40(6), 410-426. doi:10.1108/09590551211230232
- Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine learning in transportation data analytics. In M. Chowdhury, A. Apon, & K. Dey (Eds.), *Data analytics for intelligent transportation systems* (pp. 283-307). Amsterdam: Elsevier.
- Choi, T. M., Hui, C. L., Ng, S. F., & Yu, Y. (2012). Color trend forecasting of fashionable products with very few historical data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*

- (*Applications and Reviews*), 42(6), 1003-1010. doi:10.1109/TSMCC.2011.2176725
- Craparotta, G., Thomassey, S., & Biolatti, A. (2019). A siamese neural network application for sales forecasting of new fashion products using heterogeneous data. *International Journal of Computational Intelligence Systems*, 12(2), 1537-1546. doi:10.2991/ijcis.d.191122.002
- Curram, S. P., & Mingers, J. (1994). Neural networks, decision tree induction and discriminant analysis - An empirical comparison. *Journal of the Operational Research Society*, 45(4), 440-450. doi:10.1057/jors.1994.62
- Frank, C., Garg, A., Sztandera, L., & Raheja, A. (2003). Forecasting women's apparel sales using mathematical modeling. *International Journal of Clothing Science and Technology*, 15(2), 107-125. doi:10.1108/09556220310470097
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning - Data mining, inference, and prediction*. New York, NY: Springer.
- Hui, C. L., Lau, T. W., Ng, S. F., & Chan, C. C. (2005). Learning-based fuzzy colour prediction system for more effective apparel design. *International Journal of Clothing Science and Technology*, 17(5), 335-348. doi:10.1108/09556220510616192
- Hwangbo, H., Kim, E. H., & Chae, J. M. (2017). The influences of meteorological factors, discount rate, and weekend effect on the sales volume of apparel products. *Fashion & Textile Research Journal*, 19(4), 434-447. doi:10.5805/SFTI.2017.19.4.434
- Jang, E. Y., & Lee, S. J. (2002). The effects of meteorological factors on sales of apparel products - Focused on apparel sales in the department store. *Journal of the Korean Society of Costume*, 52(2), 139-150.
- Jang, E. Y., & Lim, B. H. (2003). An exploratory study on the effect of weather factors on sales of fashion apparel products in department stores. *Journal of Global Academy of Marketing Science*, 12(1), 121-134. doi:10.1080/12297119.2003.9707207
- Joo, Y. S., & Cho, G. Y. (2016). Outlier detection and treatment in industrial sampling survey. *Journal of the Korean Data & Information Science Society*, 27(1), 131-142. doi:10.7465/jkdi.2016.27.1.131
- Kim, J., & Hwangbo, H. (2017). Online and offline price elasticities of demand: Evidence from the apparel industry. *The E-Business Studies*, 18(5), 51-65. doi:10.20462/tebs.2017.10.18.5.51
- Kim, J. J. (2009). *Development of the sales forecast models of fashion products - Focusing on the case of a development stores*. Unpublished master's thesis, Hanyang University, Seoul.
- Kim, G. S., Kim, J. H., Kim, H. T., Suh, C. J., Erh, Y. Y., Yoo, S. J., Yoo, H. J., & Hwang B. J. (2011). *Production & operations management system*. Paju: Bobmunsu.
- Koo, H., & Min, D. (2013). Forecasting renewable energy using delphi survey and the economic evaluation of long-term generation mix. *Journal of the Korean Institute of Industrial Engineers*, 39(3), 183-191. doi:10.7232/JKIIE.2013.39.3.183
- Lee, E. J. (2008). *A comparative analysis of time series forecasting models for fashion products*. Unpublished master's thesis, Pukyong National University, Busan.
- Lee, K. C., & Oh, S. B. (1996). An intelligent approach to time series identification by a neural network-driven decision tree classifier. *Decision Support Systems*, 17(3), 183-197. doi:10.1016/0167-9236(95)00031-3
- Lee, S., Kang, J. H., Lee, H., Joo, T. W., Oh, S., Park, S., & Kim, S. B. (2014). Prediction of product life cycle using data mining algorithms - A case study of clothing industry. *Journal of the Korean Institute of Industrial Engineers*, 40(3), 291-298. doi:10.7232/JKIIE.2014.40.3.291
- Lee, Y. (2012). *A development study for fashion market forecasting models*. Unpublished doctoral dissertation, Ewha Womans University, Seoul.
- Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain - The bullwhip effect. *Management Science*, 43(4), 546-558. doi:10.1287/mnsc.43.4.546
- Makridakis, S. G., & Wheelwright, S. C. (1978). *Forecasting - Methods and applications*. Santa Barbara, CA: Wiley.
- Mostard, J., Teunter, R., & de Koster, R. (2011). Forecasting demand for single-period products - A case study in the apparel industry. *European Journal of Operational Research*, 211(1), 139-147. doi:10.1016/j.ejor.2010.11.001
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python - A guide for data scientists*. Sebastopol, CA: O'Reilly Media, Inc.
- Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry - A review. *Mathematical Problems in Engineering*, 2013(738675), 1-9. doi:10.1155/2013/738675
- Nam, S. M. (2006). A study on the anomaly in retailing market - Focused on the day of the week effect of sales volume in fashion apparel products retail store. *Journal of Global Academy of Marketing Science*, 16(1), 117-141. doi:10.1080/12297119.2006.9707360
- Ni, Y., & Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. *Expert Systems with Applications*, 38(3), 1529-1536. doi:10.1016/j.eswa.2010.07.065
- Saaksvuori, A., & Immonen, A. (2005). *Product lifecycle management*. New York: Springer
- Sun, Z. L., Choi, T. M., Au, K. F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1), 411-419. doi:10.1016/j.dss.2008.07.009
- Sung, H. Y. (2006). *Study of the price elasticity about the merchandises on selling in supermarkets*. Unpublished master's thesis, Chung-Ang University, Seoul.
- Sztandera, L. M., Frank, C., & Vemulapali, B. (2004). Predicting women's apparel sales by soft computing. In L. Rutkowski, J. H. Siekmann, R. Tadeusiewicz, & L. A. Zadeh (Eds.), *Artificial Intelligence and Soft Computing - ICAISC 2004: 7th International Conference, Zakopane, Poland, June 7-11, 2004. Proceedings* (pp. 1193-1198). Berlin and Heidelberg: Springer-Verlag Berlin Heidelberg.
- Thomassey, S., & Fiordaliso, A. (2006). A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems*, 42(1), 408-421. doi:10.1016/j.dss.2005.01.008
- Thomassey, S., & Happiette, M. (2007). A neural clustering and classification system for sales forecasting of new apparel items. *Applied Soft Computing*, 7(4), 1177-1187. doi:10.1016/j.asoc.2006.01.005
- Thomassey, S., Happiette, M., & Castelain, J.-M. (2003). Mean-term textile sales forecasting using families and items classification.

- Studies in Informatics and Control*, 12(1), 41-52.
- Tsujino, K., & Nishida, S. (1995). Implementation and refinement of decision trees using neural networks for hybrid knowledge acquisition. *Artificial Intelligence in Engineering*, 9(4), 265-276. doi:10.1016/0954-1810(95)00005-4
- Vashishtha, R. K., Burman, V., Kumar, R., Sethuraman, S., Sekar, A. R., & Ramanan, S. (2020, August). Product age based demand forecast model for fashion retail. *Oral presented at the fifth international workshop on fashion and KDD*. San Diego, CA.
- Vroman, P., Happiette, M., & Rabenasolo, B. (1998). Fuzzy adaptation of the holt-winter model for textile sales-forecasting. *Journal of the Textile Institute*, 89(1), 78-89. doi:10.1080/00405009808658668
- Vroman, P., Happiette, M., & Vasseur, C. (2001). A hybrid neural model for mean-term sales forecasting of textile items. *Studies in Informatics and Control*, 10(2), 149-168.
- Witten, I. H., & Frank, E. (1999). *Data mining - Practical machine learning tools and techniques with Java implementations*. San Francisco, CA: Morgan Kaufmann.
- Yelland, P. M., & Dong, X. (2014). Forecasting demand for fashion goods - A hierarchical Bayesian approach. In T. M. Choi, C. L. Hui, & Y. Yu (Eds.), *Intelligent fashion forecasting systems - Models and applications* (pp. 71-94). Heidelberg and New York: Springer.
- Yesil, E., Kaya, M., & Siradag, S. (2012). Fuzzy forecast combiner design for fast fashion demand forecasting. *Proceedings of 2012 International Symposium on Innovations in Intelligent Systems and Applications* (pp. 1-5). Trabzon, Turkey: IEEE. doi:10.1109/INISTA.2012.6247034
- Yu, Y., Choi, T. M., & Hui, C. L. (2012). An intelligent quick prediction algorithm with applications in industrial control and loading problems. *IEEE Transactions on Automation Science and Engineering*, 9(2), 276-287. doi:10.1109/TASE.2011.2173800

(Received 3 August, 2021; 1st Revised 17 August, 2021;
2nd Revised 26 August, Accepted 31 August, 2021)