

경영전문석사학위 논문

온라인 쇼핑몰 패션상품 리뷰의  
분류와 예측  
: k-평균 군집화와 LSTM을  
중심으로

2022년 2월

서울과학종합대학원대학교

김 현 숙

경영전문석사학위 논문

온라인 쇼핑몰 패션상품 리뷰의  
분류와 예측  
: k-평균 군집화와 LSTM을  
중심으로

2022년 2월

서울과학종합대학원대학교

김 현 숙

온라인 쇼핑몰 패션상품 리뷰의 분류와 예측  
: k-평균 군집화와 LSTM을 중심으로

지도교수 박 정 열

이 논문을 경영학 석사 학위논문으로 제출함

2022년 2월

서울과학종합대학원대학교

김 현 숙

김현숙의 석사 학위논문을 인준함

2022년 1월

위 원 장           고 영 희           (인)

위    원           문 달 주           (인)

위    원           박 정 열           (인)

## 초 록

패션상품의 온라인 쇼핑은 상품군별 거래액에서 높은 비중을 차지하는데 패션상품은 그 종류가 매우 다양할 뿐 아니라 개인의 선호가 차별화되어 상품기획 및 재고 관리의 효율성이 중요한 상품이다. 패션상품 구매 리뷰는 소비자의 구매에 직접적인 영향을 주며, 소비자의 선호를 파악하고 후속 구매를 예측하는 중요한 지표로 사용된다. 패션상품 리뷰 분석을 통해 소비자의 다양한 선호와 만족 불만족 요인을 밝히고 이를 적절히 활용하면 소비자 만족의 증대와 더불어 상품기획 및 마케팅 효율을 높일 수 있을 것이다. 그럼에도 불구하고 상품 리뷰는 한 문서에서 여러 주제가 동시에 존재하며, 표현 양식 및 길이가 다른 방대한 비정형 데이터로 구성되기 때문에 이를 실무에서 효율적으로 활용하기 위해서는 리뷰를 수치화한 모델링과 더불어 의미론적 분석이 함께 이루어져야 하겠다.

본 연구의 범위는 패션상품의 리뷰 데이터를 분석하여 감성예측 모델을 만들고 리뷰를 유형화하여 패션상품의 리뷰데이터 분석을 위한 감성분석 모델과 분류의 틀을 제시하는 것이다. 연구 방법은 아마존의 패션상품에 대한 리뷰 데이터셋을 사용하여 분석하였다. 아마존 패션상품 리뷰 883,663개를 분석에 사용하였고 5점 만점의 평점에 대해 1점과 2점을 부정, 5점을 긍정으로 각각 라벨링하여 인공지능 알고리즘을 사용하여 감성예측 모델과 분류 모델을 구축하고 검증하였다.

본 연구의 목적은 패션상품 리뷰를 유형화하고 감성예측 모델을 만들어 패션상품 리뷰 분류 및 예측 시스템의 기초 자료로 활용하는 것이다. 이를 위한 구체적인 연구 문제는 다음과 같다. 첫째, 온라인 쇼핑몰의 패션상품 리뷰의 감성예측 모델을 만들고 성능을 검증한다. 둘째, 온라인 쇼핑몰의 패션상품 리뷰를 분류하고 성능을 검증한다.

패션상품 온라인 리뷰의 감성예측 모델을 구축하기 위해 다양한 임베딩과 알고리즘이 결합된 모델을 검증하였다. 임베딩으로 Count vectori-

zation, TF-IDF, Word2Vec을 사용하였고, 알고리즘은 SVC, logistic regression, random forest classifier, bagging classifier, multinomial NB을 사용하여 검증하였다. 이들 결과를 bi-directional sequence LSTM 모델과 비교한 결과 bi-directional LSTM 모델의 성능이 정확도 0.94로 가장 높게 나타났다. 또 패션상품 리뷰를 k-평균 군집화 알고리즘으로 유형화한 결과 3개 집단으로 분류되었고, 각 집단을 “Size & Fit”, “Quality & Price”, “Appearance”로 명명한 다음 ANOVA 평균분석 및 사후분석과 실루엣스코어로 유의성을 검증하였다. 또 bi-directional LSTM을 분류된 세 집단에 대해 테스트한 결과 세 집단 모두 정확도가 0.93으로 나왔다. 패션상품 리뷰를 3개 유형으로 분류한 모델의 성능을 테스트하기 위해 데이터셋을 구성하는 각 포인트 별로 리뷰 내용을 대표할 수 있는 한줄 총평을 세 가지 분류기준으로 라벨링하여 bi-directional LSTM으로 모델링한 결과 정확도가 0.79로 나왔다.

결론적으로 본 연구는 패션상품 리뷰의 분류와 감성 예측 시스템 구축을 위한 모델을 제안함으로써 소비자의 구매 후 만족을 결정하는 요인을 파악하고, 리뷰 기반의 효율적인 상품 기획 및 마케팅 전략에 필요한 기초자료를 제공하였다. 또 본 연구의 평점 예측 모델은 평점이 없는 리뷰 텍스트의 긍·부정 예측에 효율적으로 사용될 수 있을 것이며, 가짜 평점을 분류해 상위 노출을 방지하는 등 리뷰의 효율적 관리에 사용될 수 있을 것이다. 또한 패션상품 리뷰 텍스트의 분류 기준을 적용하여 리뷰를 카테고리별로 관리한다면 소비자의 요구에 대한 보다 효율적이고 신속한 대응이 가능할 것이다.

주제어: 온라인 리뷰, 패션 상품, 유형화, k-평균 군집화, LSTM, 감성분석

# 목 차

제 I 장 서론 .....	1
제 1 절 연구의 배경 및 목적 .....	1
(1) 연구 필요성 .....	1
(2) 연구 목적 .....	3
제 2 절 연구의 내용 및 방법 .....	5
(1) 연구 내용 및 범위 .....	5
(2) 연구 방법 및 절차 .....	5
제 3 절 연구의 기대효과와 활용 방안 .....	7
제 II 장 이론적 배경 .....	8
제 1 절 구매 리뷰 .....	8
(1) 온라인 쇼핑몰 구매 리뷰 .....	8
(2) 패션상품 구매 리뷰 .....	11
제 2 절 구매 리뷰 예측 .....	14
(1) 구매 리뷰와 감성 분석 .....	14
(2) 감성 예측 모델 .....	15
제 3 절 구매 리뷰 분류 .....	20
(1) 구매 리뷰 유형화 .....	20
(2) 토픽분석 .....	21
(2) 유형화 모델 .....	22
제 III 장 분류 및 예측모델 개발 .....	23
제 1 절 데이터 수집 및 전처리 .....	23

(1) 데이터 수집 .....	23
(2) 전처리 .....	23
제 2 절 알고리즘 구축 .....	32
(1) 데이터 임베딩 .....	32
(2) 학습 알고리즘 .....	32
제 3 절 예측 시뮬레이션 .....	34
(1) 감성 예측 .....	34
(2) 성능 평가 .....	35
(3) 감성별 토픽 .....	40
제 4 절 유형 분류 .....	47
(1) 리뷰 유형 분류 .....	47
(2) 성능 평가 .....	51
<b>제 IV 장 결론 .....</b>	<b>55</b>
제 1 절 요약 및 시사점 .....	55
제 2 절 연구의 한계 및 제언 .....	58
<b>참고문헌 .....</b>	<b>59</b>
<b>Abstract .....</b>	<b>64</b>

## 표 목 차

<표 1> 리뷰 텍스트의 polarity 분포 .....	27
<표 2> 긍정 텍스트 리뷰 샘플 .....	28
<표 3> 중립 텍스트 리뷰 샘플 .....	28
<표 4> 부정 텍스트 리뷰 샘플 .....	29
<표 5> 데이터 임베딩과 학습 알고리즘 테스트 .....	33
<표 6> 모델 정확도 .....	35
<표 7> CNN 모델 .....	36
<표 8> CNN 모델 파라미터 .....	36
<표 9> LSTM 모델 .....	38
<표 10> LSTM 모델 파라미터 .....	38
<표 11> 리뷰 감성 예측 샘플 .....	39
<표 12> 긍정 리뷰의 토픽 .....	41
<표 13> 부정 리뷰의 토픽 .....	44
<표 14> 리뷰 유형별 n-gram .....	49
<표 15> 집단별 평점 평균 .....	53
<표 16> ANOVA .....	53
<표 17> Turkey HSD 평균 다중 비교 .....	53



## 그림 목 차

<그림 1> 연구 절차 .....	6
<그림 2> 아마존 리뷰 예시 .....	23
<그림 3> 데이터셋의 평점별 분류 .....	24
<그림 4> 리뷰 텍스트와 총평의 polarity 분포 .....	25
<그림 5> 평점별 리뷰 텍스트의 polarity 분포 .....	26
<그림 6> 텍스트 리뷰 polarity의 긍정, 부정, 중립 분포 .....	26
<그림 7> 에포크에 따른 훈련데이터와 테스트데이터 정확도 .....	27
<그림 8> 리뷰 텍스트의 top-1gram .....	30
<그림 9> 리뷰 텍스트의 top-2gram .....	31
<그림 10> 리뷰 텍스트의 top-3gram .....	31
<그림 11> 에포크에 따른 정확도 .....	37
<그림 12> LSA에 의한 긍정 리뷰 토픽 .....	42
<그림 13> LDA에 의한 긍정 리뷰 토픽 .....	42
<그림 14> LSA에 의한 긍정 리뷰의 t-SNE .....	43
<그림 15> LDA에 의한 긍정 리뷰의 t-SNE .....	43
<그림 16> LSA에 의한 부정 리뷰 토픽 .....	45
<그림 17> LDA에 의한 부정 리뷰 토픽 .....	45
<그림 18> LSA에 의한 부정 리뷰의 t-SNE .....	46
<그림 19> LDA에 의한 부정 리뷰의 t-SNE .....	46
<그림 20> 군집 수에 따른 K-평균 군집화 결과 .....	47
<그림 21> 데이터의 3-D점 표현 .....	48
<그림 22> DBSCAN 결과 .....	48
<그림 23> ‘Size & Fitting’ 그룹의 n-gram .....	49
<그림 24> ‘Price & Quality’ 그룹의 n-gram .....	50

<그림 25> ‘Styling’ 그룹의 n-gram .....	50
<그림 26> 실루엣 점수 .....	52
<그림 27> 집단 평균 .....	53

# 제 I 장 서 론

## 제1절 연구의 배경 및 목적

### (1) 연구 필요성

온라인 리뷰는 소비자의 구매 결정에 영향을 주므로 판매자나 구매자 모두에게 중요한 의미를 가진다. 소비자의 상품에 대한 리뷰 데이터 분석을 통해 소비자의 피드백을 반영하는 것은 소비자 만족을 위해서 뿐 아니라 기업의 효율적인 경영성과 창출에 필요하다(McKinney & Shim, 2016). 온라인 쇼핑에서 패션상품은 상품군별 거래액에서 높은 비중을 차지하는 품목으로 상품구성이 다양하고 같은 상품이라도 소비자에 따라 선호 및 맞춤새가 다르다. 이같이 패션상품은 소비자 구매 만족에 개인 및 상황 변수가 크게 작용하는 상품임에도 불구하고 상품을 직접 입어보고 구매할 수 없다는 온라인 쇼핑의 특성 때문에 타인의 구매 리뷰가 구매의 사결정에 미치는 영향이 크다(이동엽 외, 2017).

상품구매 리뷰는 소비자의 선호를 파악하고 후속 구매를 예측하는데 중요한 지표임에도 불구하고 리뷰는 주관적으로 작성되는 것이기 때문에 객관적으로 해당 상품에 대한 소비자 성향을 파악하는 것이 어렵다. 또한 구매 리뷰는 하나의 리뷰에 여러 주제가 동시에 언급되며 표현 양식 및 길이가 다른 방대한 양의 비정형 데이터로 구성되기 때문에 실무에서 활용하는데 어려움이 있다(Almiron-chamadoira, 2018)..

특히 패션상품은 종류가 매우 다양하며 선호의 개인차가 크고 계절성이 강해 재고 관리가 중요한 상품이다. 패션상품에 대한 소비자의 만족과 더불어 효율적인 재고 관리와 상품 판매 증가를 위해서는 온라인 상품구매 리뷰를 적절히 분석하여 활용할 필요가 있다. 패션상품의 구매 리뷰로부터 다른 상품군과 구별되는 패션상품 고유의 만족 요인 및 구매 결정 요인을 파악할

필요가 있으며 이를 위해서는 만족 불만족을 수치화하여 예측하는 양적 분석뿐 아니라 리뷰 텍스트에서 언급된 내용에 관한 의미론적 분석이 함께 이루어져야 할 것이다(한기향, 2021).

온라인 리뷰에 관한 연구가 활발히 이루어지고 있는데 대표적인 연구로 온라인 리뷰의 분석을 통해 평점이나 감성을 예측한 연구, 소비자 행동이나 매출에 리뷰가 미치는 영향에 대한 연구 등이 있다(Chong et al., 2016; Elmutngi & Gherbi, 2008; Kim & Kim, 2018; Lee et. al., 2020; 이정원, 박철, 2020). 선행연구에서는 온라인 리뷰의 내용뿐 아니라 리뷰의 특성, 길이, 리뷰 수, 리뷰 작성자 등을 독립변수로 하여 이들 변수가 소비자 감성이나 평점, 제품 및 서비스의 흥행, 매출, 인지도나 구매의도 등에 미치는 영향을 분석하였다(Reich & Maglio, 2020; Chang & Liu, 2019; 이동엽 외, 2017; 이윤주 외, 2020; 이정원, 박철, 2020; 한기향, 2021). 리뷰로부터 주요 마케팅 변수를 예측하기 위한 모델링이나 내용분석이 이루어졌는데 대부분 리뷰의 내용에 대한 분석과 예측을 위한 모델링이 따로 이루어져 연구결과를 리뷰 분석 시스템에 활용되는데 제약이 있다. 또 패션상품의 카테고리에 적합한 리뷰 분류가 다양하게 시도되고 있기는 하나 리뷰 자체가 단일 내용으로 구성되어 있지 않기 때문에 그 분류가 힘들고 분류의 재현성을 달성하기 힘든 측면이 있다. 쇼핑물은 리뷰의 영역을 세분화하여 데이터를 수집할 때 마케팅에 활용 가능성이 높아질 뿐 아니라 쇼핑물 사이트에서 세분 영역별로 리뷰를 분류하여 제시함으로써 소비자의 검색과 필터링을 용이하게 해줄 수 있다. 따라서 패션상품에 특화된 리뷰 분류 체계를 제안하고 리뷰의 효율적인 관리는 위한 모델을 구축하는 보다 확장된 연구가 필요하다.

한편 리뷰와 평점의 관계에서 소비자마다 만족과 불만을 느끼는 요인에 차이가 있어 평점 평균만으로 정확히 소비자 만족 불만족의 상태를 파악하기 힘들다(김준겸 외, 2021). 또 소비자들이 평점과 리뷰 내용을 상이하게 입력하는 경우가 있어 리뷰 신뢰도와 유용성에 대한 의문이 제기되고 있다. 간혹 불공정한 가짜 구매 리뷰가 있어 이를 탐지하여 가능한 빨리

리 제거하는 것이 필요하다(Want & Chen, 2020). 이같이 평점에만 의지해 소비자 만족도를 판단하거나 상품 정책을 전개하는 것은 위험하며 평점과 리뷰 내용의 일치도를 평가하여 효율적으로 대응해야 한다. 이를 위해 AI를 활용한 모델링으로 상시적으로 리뷰를 관리할 수 있는 시스템에 대한 연구가 필요하다. 즉 온라인 리뷰가 상품기획, 마케팅 전략 및 고객 서비스에서 중요한 역할을 함에도 불구하고 온라인 리뷰를 효율적으로 반영하려면 수적 모델링이나 내용분석에서 나아가 보다 세분화된 연구가 필요하다.

## (2) 연구 목적

온라인 상품 리뷰를 마케팅에 적절히 활용하고 고객 불만족에 신속하게 대처하여 소비자 만족을 극대화하기 위해서는 빅데이터 분석을 위한 머신러닝 알고리즘을 활용하고 다양한 질적, 양적 연구 방법을 결합하여 온라인 리뷰를 다면적으로 분석할 필요가 있다. 특히 패션상품의 구매 리뷰를 효율적으로 활용하려면 패션상품 고유의 특성을 반영한 분석 체계가 이루어져야 하겠다. 이에 본 연구에서는 패션상품 리뷰의 분류 체계를 만들고 감성 예측 시스템을 검증하여 향후 리뷰 분석의 효율성을 높이고자 한다.

본 연구의 목적은 온라인 패션몰 상품구매 리뷰를 유형화하여 토픽을 분석하고 각 유형별로 감성 예측 모델을 만들어 성능을 비교하는 것이다. 더불어 각 분석 단계별로 서로 다른 임베딩 방법과 알고리즘의 차이를 검증할 것이다. 이를 위한 구체적인 연구문제는 다음과 같다.

연구문제 1. 패션상품 구매 리뷰로부터 감성을 예측하고 평가한다.

- 1-1. 패션상품 구매 리뷰의 감성을 예측한다.
- 1-2. 패션상품 구매 리뷰 감성 예측 모델을 평가한다.
- 1-3. 긍·부정 감성의 토픽을 분석한다.

연구문제 2. 쇼핑 패션상품 구매 리뷰를 분류하고 검증한다.

- 2-1. 패션상품 구매 리뷰를 분류한다.

2-2. 패션상품 구매 리뷰 분류의 정확도를 검증한다.

## 제2절 연구의 내용 및 방법

### (1) 연구의 내용 및 범위

본 연구는 온라인 쇼핑몰 패션 구매상품 리뷰를 분석하기 위해 아마존의 패션상품 리뷰 데이터셋을 사용하였다. 아마존 패션 리뷰 데이터로부터 긍·부정 평점을 예측하기 위한 감성예측 모델을 개발하고 텍스트를 유형화하여 각 유형 별로 텍스트를 대표하는 벡터가 평점을 예측하는 정도와 리뷰 대신 총평을 넣었을 때의 분류성과를 분석하였다.

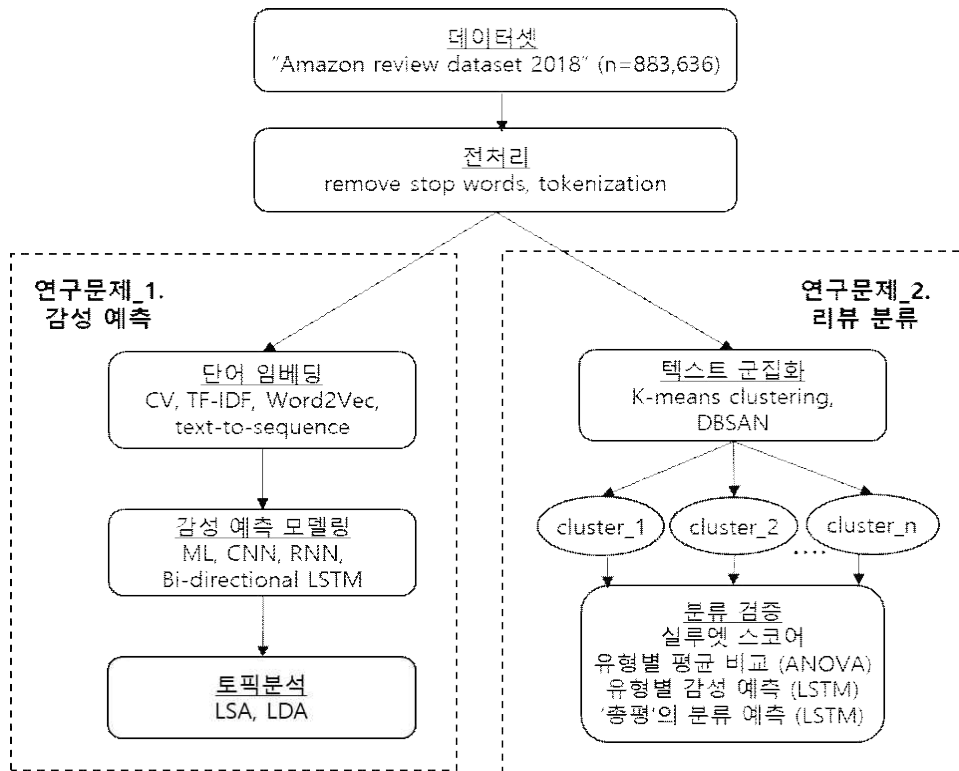
본 연구의 범위는 패션상품의 리뷰 데이터를 분석하여 감성을 예측하고 리뷰를 유형화하여 분류 예측모델을 제시하는 것으로 리뷰 분석을 위한 실무적, 이론적 기초 자료를 제시하고자 한다.

### (2) 연구 방법 및 절차

연구 방법은 아마존의 패션상품 리뷰 데이터셋을 사용하여 <그림 1>의 연구 절차와 같이 텍스트 데이터를 분석하였다. 아마존 패션상품 리뷰 883,663개를 분석에 사용하였고 5점 만점의 평점에 대해 1점과 2점을 부정, 5점을 긍정으로 각각 라벨링하여 인공신경망분석을 사용하여 감성예측 모델을 검증하였다. <그림 1>의 연구문제 1에서 단어 임베딩 방법으로 CV, IF-IDF, Word2Vec, text-to-sequence를 사용하였고, 감성 예측 모델링 방법으로 ML에서 SVC, logistic regression, random forest classifier, bagging classifier, multinomial NB를 사용하였고, 딥러닝 모델로 CNN, RNN, Bidirectional LSTM 모델을 테스트하였다. 특별히 LSTM은 Agarap(2020)의 연구에서 성분이 사라지는 RNN의 단점을 보완한 감성 분석 알고리즘으로 제안하였으며 리뷰 텍스트에서 단어의 문맥을 보다 잘 파악하기 위해 쓰인 LSTM 알고리즘은 좋은 성능을 나타냈다. 한편 리뷰 텍스트의 긍·부정 리뷰의 컨텐츠 내용을 설명하기 위해 LSA, LDA로 토픽분석을 하여 주제어를 추출하였다.

<그림 1> 연구문제 2에서는 리뷰의 유형을 분류하기 위해 k-평균 군

집화와 DBSCAN을 사용하여 분류하였고 그 결과의 유효성을 검증하기 위해 실루엣 스코어, ANOVA 검증 등을 사용하였다. 또 분류된 집단별로 평점예측 모델의 감성 예측 정확도를 검증하고, 리뷰 대신 총평을 넣어 분류 성과를 살펴보았다. 리뷰 분류 시스템을 총평으로 다시 검증하는 것은 Praveenraj, et al.,(2021)가 남성 패션 액세서리 구입의 온라인 구매 리뷰를 대상으로 감성 분석을 한 연구에서 코멘트의 감성 점수와 타이틀의 감성 점수에서 평점과 감성 점수가 관련이 있다고 한 연구결과에 기초한다. 마지막으로 감성예측 모델 및 패션상품 리뷰 분류 모델의 실무 적용 가능성 및 이론적 기여에 대해 논하였다.



<그림 1> 연구 절차



### 제3절 연구의 기대효과와 활용 방안

본 연구는 온라인 쇼핑몰의 패션상품 구매 리뷰로부터 빅데이터 분석을 통해 감성 예측모델을 만들고 리뷰의 내용을 유형화하여 의미론적으로 분석하고자 한다. 감성 예측 및 분류의 틀을 제시함으로써 향후 해당 분야의 연구에 필요한 기초자료로 사용될 수 있을 것이고 실무에서 효율적인 리뷰 관리 및 사용에 도움이 되리라 기대할 수 있다.

온라인 마케팅 실무에서 리뷰를 세분화된 영역별로 활용할 때 보다 효율성이 높아질 것이며 소비자에게 세분 영역별로 리뷰를 제시함으로써 검색과 필터링을 쉽게 해줄 수 있을 것이다. 또한 온라인 상품구매 리뷰에 다양한 내용이 섞여 있는 경우가 많아 자세히 읽어 보기 전에는 대표적인 내용을 추출하거나 긍·부정 감성을 파악하기 힘들며 때때로 거짓 평점으로 인해 리뷰와 평점이 일치하지 않는 경우가 있다. 이같이 온라인 상품구매 리뷰 데이터가 빅데이터임에도 불구하고 효율적으로 사용되기 힘들다는 점을 고려할 때 본 연구에서 정확도가 높은 최적의 알고리즘을 구축한다면 향후 구매 리뷰의 감성 예측 시스템을 운영하여 빠른 시간에 상품 판매를 예측하여 재고관리에 활용하고, 고객 불만족에 신속하게 대응할 수 있을 것이다. 또한 평점이 없는 경우의 블로그나 SNS 글에 대해서도 감성을 판별할 수 있으며, 거짓 리뷰를 판별하여 대응할 수 있을 것이다. 결론적으로 본 연구는 패션 리뷰 연구에 기초자료로 활용될 수 있을 것이며 온라인 마케팅 실무와 고객관리, 리뷰 관리에 효율적으로 활용될 것이다.

## 제 II 장 이론적 배경

### 제1절 구매 리뷰

#### (1) 온라인 쇼핑몰 구매 리뷰

온라인 쇼핑몰 제품구매 리뷰는 온라인 구전의 대표적인 형태로 소비자가 제품 구매 후 남긴 글을 말한다(이호근, 곽현, 2013). 구매 리뷰는 소비자의 제품 구매 의사 결정에 영향을 줄 수 있으므로 소비자에게나 판매자 모두에게 중요한 의미를 가진다(Almiron-Chamadoira, 2018). 온라인 리뷰의 실무적 중요성에 따라 연구가 활발히 이루어지고 있는데 대표적인 연구로 온라인 리뷰의 내용을 분석하거나, 리뷰로부터 평점이나 감성을 예측한 연구, 혹은 리뷰가 소비자 행동이나 매출에 미치는 영향에 대한 연구 등이 있다(Chong et al., 2016; Elmutngi & Gherbi, 2008; Kim & Kim, 2018; Lee et. al., 2020; 이정원, 박철, 2020). 또 온라인 리뷰의 내용뿐 아니라 리뷰의 특성, 길이, 리뷰 수, 리뷰 작성자 등을 독립변수로 하여 이들 변수가 소비자 감성이나 평점, 제품 및 서비스의 흥행, 매출, 인지도나 구매의도 등에 미치는 영향에 대한 연구가 있다(Reich & Maglio, 2020; Chang & Liu, 2019; 이동엽 외, 2017; 이윤주 외, 2020; 이정원, 박철, 2020; 한기향, 2021).

리뷰의 수, 길이, 방향성과 같은 리뷰의 특성이나 리뷰 작성자의 특성이 매출 성과에 미치는 영향에 대한 연구가 활발히 이루어졌다. 리뷰의 방향성의 효과에 대한 연구로 이정원, 박철(2020)의 연구에서는 온라인 구전량과 온라인 구전 방향성이 매출액에 긍정적인 영향을 미치는 것으로 나왔으며, 온라인 구전의 분산은 영화 매출액에 부정적인 영향을 미치는 것으로 나타났다. 특히 온라인 구전 방향성과 구전 분산 효과는 미국에 비해 한국에서 긍정적 효과가 강하다. 온라인 구전의 방향성은 제품의 품질 등 지각된 위험을 감소하는 역할을 하기 때문에, 불확실성 회피의 효과있음이 증명 되었다.

리뷰의 수가 미치는 영향에 대한 연구로 Duan, et al. (2008)은 온라인 리뷰가 영화의 박스오피스 성과에 미치는 설득효과와 인지효과를 검증하였다. 온라인 리뷰는 영화 매출에 영향을 주며 영향을 받기도 하는 것으로 나타났으며, 온라인 리뷰는 그 자체가 구매결정에 영향을 주기 보다는 리뷰의 숫자가 인지도에 영향을 주는 것으로 나타났다. 리뷰의 방향성과 양이 미치는 영향에 대한 연구로 Chong et al. (2016)은 감성분석과 인공지능망 분석을 통해 온라인 리뷰의 방향성(valence)과 양(volume), 온라인 홍보 전략 (무료배송과 할인) 및 리뷰의 감성이 제품 판매에 미치는 영향을 MLP(multilayer perceptron) 모델로 예측하였다. 그 결과 온라인 리뷰, 온라인 홍보전략, 온라인 감성이 제품 매출에 미치는 영향이 각자 달랐으며 상호작용 효과가 개별 효과보다 중요한 것으로 나왔다. 특별히 온라인 리뷰 수는 감성이나 할인과 상호작용 효과가 컸다. 온라인 평점이 판매에 미치는 영향은 유의하지 않았다. 리뷰의 길이에 대한 연구로 조용희 외(2018)의 연구에서는 영화를 흥행 성공, 부진, 실패로 구분하고 영화 평점과 리뷰 길이와의 상관관계 및 회귀분석을 하였다. 그 결과 영화에 대해 불만족스러운 경우 즉 평점이 낮을 때 만족스러운 경우보다 영화 리뷰의 길이가 길었으며 이는 불만족한 사람이 리뷰를 쓸 때 왜 불만족했는지 더 구체적으로 설명함을 알 수 있다.

리뷰 작성자의 특성이 소비자 구매에 미치는 영향에 대한 연구로 Reich and Maglio(2020)은 “Sephora” 쇼핑몰에서 이전에 잘못 구매한 경험이 있는 리뷰어가 추천하는 제품을 그렇지 않은 리뷰어가 추천하는 제품보다 더 많이 구매했다고 하였다. Zhang and Liu(2019)은 리뷰어의 소셜네트워크가 온라인 리뷰에 미치는 영향을 분석하기 위해 인플루언서의 개인 네트워크와 리뷰를 분석한 결과 여러 네트워크 연결이 있는 리뷰어는 리뷰의 긍정성을 오히려 감소시켜 미래의 리뷰평가에 부정적 영향을 미치는 것으로 나왔다.

한편 리뷰와 평점의 관계에서 소비자마다 만족과 불만을 느끼는 요인에 차이가 있어 평점 평균만으로 정확히 소비자 만족 불만족의 상태를 파악하기 힘들다(김준겸 외, 2021). 그런데 부정적인 온라인 리뷰에 대한 피드백을 적절히 제공하지 못하면 소비자 신뢰와 명성에 문제가 생기게 된다. 그래서

Elmurngi and Gherbi(2018)는 부정적인 온라인 리뷰에 대해 연구하였는데 온라인 리뷰의 감성 분석과 관련되어 중요한 것은 어떻게 불공정하게 부정적이거나 긍정적이거나, 중립인 리뷰를 감지하느냐 하는 것이라 하였다. 그들은 의복, 신발, 주얼리 리뷰, 유아용품 리뷰, 애완동물 물품 리뷰의 세 개의 데이터 셋에 대해 감성분석을 한 결과 불공정한 리뷰(unfair review)를 분석하는데 효율적인 머신러닝 알고리즘으로 로지스틱 회귀분석이 가장 효과적인 것으로 나타났다.

소비자들이 평점과 리뷰 내용을 상이하게 입력하는 경우가 있어 리뷰의 신뢰도와 유용성에 대한 의문이 제기되고 있다. 간혹 불공정한 가짜 구매 리뷰가 있어 이를 탐지하여 가능한 빨리 제거하는 것이 필요하다 (Want & Chen, 2020). 김준겸 외(2021)는 온라인 리뷰와 평점이 일치하지 않는 경우의 분석을 위해 리뷰의 세분 영역에서의 주요 키워드를 품사별로 추출하고 부정어 사전을 만들었다.

온라인 리뷰에 대한 연구방법은 크게 수치에 의한 모델링 연구와 리뷰의 내용에 중점을 둔 의미론적 분석이 있다. 리뷰의 내용에 중점을 둔 연구로 이윤주 외(2020)는 흥행의 지표를 웹툰 플랫폼 정식연재로 보고 성공한 웹툰과 그렇지 못한 웹툰 14개의 댓글 36,194개를 TF-IDF로 주요 단어를 비교 분석하였다. 또 Kim et al.(2018)은 신발에 대한 온라인 리뷰를 내용 분석하고, 구매의 시간적 거리가 온라인 리뷰의 조절초점이 예방초점이나 향상초점이나에 따라 구매자 태도와 구매의도에 다르게 영향을 주는지 실험 연구를 하였다. 그 결과 가까운 미래일 때에는 예방초점의 리뷰에 대해 보다 호의적인 태도를 보였고, 먼 미래의 구매에 대해서는 조절초점에 따른 차이가 나타나지 않았다.

위에서 본 바와 같이 선행연구에서 리뷰로부터 주요 마케팅 변수를 예측하기 위한 모델링이나 내용분석이 활발히 이루어졌는데 연구결과를 실무에서 효율적으로 적용하기 위해서는 수치에 의한 모델링과 내용에 중점을 둔 의미론적 분석을 통합한 모델을 제시할 필요가 있다. 평점에만 의지해 소비자 만족도를 판단하거나 상품 정책을 전개하는 것은 위험하며 평점과 리뷰 내용의 일치도를 평가하여 효율적으로 대응해야 한다. 이를 위해 AI를 활용한 모델링으로 상시적으로 리뷰를 관리할 수 있는 시스템

에 대한 연구가 필요하며 온라인 리뷰가 상품기획, 마케팅 전략 및 고객 서비스에서 효율적으로 반영되려면 수적 모델링이나 내용분석에서 나아가 보다 세분화되고 통합된 연구가 필요하다.

## (2) 패션상품 구매 리뷰

온라인 리뷰는 상품이나 서비스 카테고리에 따라 다른 성격을 가지므로 이를 고려한 연구가 필요하다. Almiron-Chamadoira(2018)은 온라인 리뷰가 상품의 카테고리 장르에 따라 공통되는 일반적인 내용과 상품 특이적인 부분으로 구성된다고 하고 상품 카테고리 별로 온라인 리뷰 구성 요소를 연구한 결과 온라인 리뷰는 장점, 품질, 단점, 그 외의 내용으로 구성된다고 하였다. 온라인 리뷰의 효과를 정확히 예측하기 위해서는 상품의 카테고리 별로 고유한 부분을 규명하여 상품의 특성을 반영한 분석이 필요하다. 여러 연구에서 특정 제품이나 서비스 군을 대상으로 온라인 리뷰를 유형화하고 고객행동을 분석하였는데 Lee et al. (2020)은 에어비앤비의 온라인 리뷰를 통해 고객 행동을 분석하여 계층적 군집분석 알고리즘으로 리뷰를 유형화하고 좋은 숙소에 대한 에어비앤비 사용자의 마인드셋을 파악하였다.

패션상품 구매리뷰에 대한 선행연구에는 리뷰의 특성에 대한 연구, 리뷰의 내용 분석, 감성 분석 예측 모델 연구 및 리뷰의 유형화 연구 등이 있다. 각 주제 별 선행연구를 살펴보면 다음과 같다. **패션상품 구매리뷰의 특성에 관한 연구**로 Kawarf & Istanbuluoglu, (2019)은 온라인 쇼핑몰의 구매 리뷰와 페이스북의 글을 인터뷰를 통해 비교하였다. 그 결과 패션 영역에서 고객 리뷰와 페이스북 활동의 특성이 다르다는 것을 발견하였다. 패션은 정체성이나 자기표현과 관련이 깊기 때문에 패션 온라인 고객 리뷰는 유동성을 가지는데 이러한 유동성을 개념화하는 틀이 필요하다. 패션 고객은 사이즈나 색상의 선택에 리뷰의 영향을 받는데 선호에는 큰 영향을 받지 않아 리뷰가 구매를 촉진하는 요인이기 보다는 불만족 요인을 체크하는 정도의 역할을 하는 것으로 볼

수 있겠다. 이렇게 제품이나 서비스의 종류에 따라 리뷰가 구매에 미치는 영향이 다름에 따라 패션 구매 리뷰를 독립적으로 세분화하여 연구할 필요가 있으며 패션 리뷰의 유동성을 반영하면서 리뷰의 분류 틀을 설정하는 연구가 필요하다. 리뷰의 특성에 대한 연구로 주보라와 황선진(2016)은 부정적 리뷰를 조작하여 부정적 리뷰가 패션 제품 선호에 미치는 영향을 실험 연구로 분석하였다. 그 결과 부정적 리뷰의 유형이 주관적일 때 보다 객관적일 때 소비자 제품 선호에 더 큰 영향을 주는 것으로 나타났다.

패션구매 리뷰가 유동적이고 비체계적임에 따라 선행연구에서 패션 리뷰의 유형화를 시도하였다. 패션 리뷰의 유형화 연구에는 리뷰를 직접 분류하여 코딩한 연구(McKinney & Shim, 2016), 이론적으로 긍정적 리뷰와 부정적 리뷰, 유용적 리뷰와 쾌락적 리뷰로 분류하여 분석한 연구(채희주 외, 2016) 등이 있다. McKinney and Shim(2016)의 연구에서는 정장 렌탈 경험에 대한 리뷰를 내용분석하여 정장 렌탈의 평가 기준을 세가지 유형으로 도출하였다. 정장 렌탈 평가 기준은 맞춤새, 의복 스타일, 사회적 반응, 적합성, 칼라와 패턴, 소재, 신체적 편안함으로 구성되는 의복평가기준과, 전통적 서비스 품질, 인터넷의 서비스 품질로 구성되는 e-서비스 평가 기준, 마지막으로 보다 나은 핏과 외모를 위한 전략, 자기 서술적 정보, 추천, 미래의 구매 의향 등의 기타 내용으로 구성되었다. 한편 채희주 외(2016)의 연구에서는 패션 상품 리뷰를 긍·부정 리뷰, 도구적·쾌락적 리뷰로 분류하여 분석한 결과 리뷰가 도구적이면서 부정적인 내용일 때 소비자의 신뢰 형성에 가장 큰 영향을 주었다. 이처럼 패션 리뷰는 다차원적인 속성이 있으므로 유형화하고 분류하여 세분화된 영역에 따라 다른 전략을 구사할 필요가 있다.

패션 구매 리뷰의 내용이 중요함에 따라 리뷰의 감성 및 토픽을 분석한 연구들이 많다. 한기향(2021)은 아마존에서 크롤링한 반팔 티셔츠에 대한 구매 리뷰를 형태소 분석, 빈도분석, TF-IDF, 감성분석, 연결중심성으로 분석한 결과 반팔 티셔츠 구매시 가장 중요하게 생각하는 요인이 ‘사이즈’임을 알 수 있었

다. 사이즈 다음으로 심미적 요인, 소재의 품질, 세탁, 활용성이 중요하게 나타났다. 또 감성 분석 결과 긍정의 비율이 높으며 심미적 요인과 연관된 예쁘다, 귀엽다가 많이 나왔고 사이즈에 대한 긍정의 댓글도 많았다. 패션 리뷰에 머신러닝 알고리즘을 사용하여 감성을 분석한 연구로 Elmurngi and Gherbi(2018)은 의복, 신발, 주얼리 리뷰, 유아용품 리뷰, 애완동물 물품 리뷰의 세 개의 데이터 셋에 대해 4가지 머신러닝 알고리즘 (Naive Bayes, Decision Tree, Logistic Regression, Support Vector Machine)을 사용하여 감성 분석을 하고 모델 성능을 비교하였다. 그 결과 로지스틱 회귀분석이 텍스트 분류 뿐 아니라 불공정한 리뷰를 감지하는데 효과적이었다.

패션상품의 구매 리뷰에 대한 선행 연구를 살펴본 결과 패션 상품 리뷰 고유의 특성을 파악하려는 연구가 많았고 이와 더불어 패션 상품의 리뷰가 다면적이고 유동적임에 따라 그 내용을 유형화하고 분류하려는 시도가 있었다. 한편 패션 구매 리뷰에 인공지능 모델을 활용한 연구는 상대적으로 많지 않았다. 구매 리뷰가 방대한 양의 빅데이터로 쌓임에 따라 이를 효율적으로 관리하기 위해서는 인공지능 기반의 모델링 연구가 유형화 연구와 더불어 보다 활성화될 필요가 있다.

## 제2절 구매 리뷰 예측

제2절에서는 구매 리뷰의 예측을 위한 감성 분석 연구를 살펴보고, 감성 예측을 위한 텍스트 마이닝에 필요한 임베딩 방법 및 알고리즘에 대한 선행연구를 살펴보겠다.

### (1) 구매 리뷰와 감성 분석

감성분석(sentiment analysis)이란 텍스트에서 의견, 정서, 감정 및 태도를 분석하여 감성지수를 계산하고, 긍정지수, 부정지수로 구성된 감성지수를 사용하여 텍스트에 포함된 긍정, 부정의 감성을 판단하는 기법이다(윤소영, 윤성대, 2020). 온라인 구매가 활발히 일어나고 있는 현 시점에서 방대한 양의 사용자 리뷰에 대해 감성분석을 통해 사용자 만족을 효과적으로 증가시킬 수 있을 것이다(Yang, et. al, 2020).

온라인 리뷰에 감성분석을 적용한 연구에는 긍·부정의 감성에 해당하는 핵심 단어를 추출하는 연구, 인공신경망 분석을 통해 감성을 예측하는 모델링 연구 등이 있으며 감성분석 결과를 활용하여 다른 변수를 예측하거나 다른 변수와 결합시켜 알고리즘을 향상시키는 등의 연구가 있다(윤소영, 윤성대, 2020; 이동엽 외, 2017; 이지현 외, 2020; 최영현, 이규혜, 2020; Chong et al., 2016; Elmurngi, & Gherbi, 2018; Yang, et al., 2000). 각각의 주제에 해당하는 연구를 아래에서 살펴 보겠다.

온라인 패션 구매 리뷰에서 긍·부정의 감성에 해당하는 핵심 단어를 추출한 최영현, 이규혜(2020)의 연구에서는 인터넷 카페, 네이버, 다음, 유튜브에서 제공하는 웹문서를 대상으로 패션제품 리뷰에서 빈도와 TF-IDF로 핵심 단어를 추출하여 감성 사전을 구축하고 베이지안 분류기(Bayes classifier)로 긍정·중립·부정의 감성을 예측하는 모델을 만들었다. 그 결과 패션 리뷰에서 구매 결정 요소로 가격, 크기, 장식, 배송, 세탁 및 소매상의 안정성, 제품구매 동기가 추출되었고, 해결해야 할 불만족 요소로 불충분한 양, 공급 배송지연, 고가가격, 불량한 장식, 퇴색한 색상, 세탁물 오염, 구매 후 불쾌한 냄새가 나왔다.



감성예측 모델링 연구로 이동엽 외(2017)의 연구에서는 아마존 패션상품에 대한 리뷰 570만건을 Word2Vec으로 임베딩하고 SVM분류기 모델로 학습하여 사용자 감성을 예측하였고 그 결과 88.0%의 정확도를 얻었다. 한편 감성분석 모델의 알고리즘을 개선하기 위한 연구로 Yang, et al.(2020)은 감성 어휘와 CNN, BiGRU (attention-based Bidirectional Gated Recurrent Unit)를 결합하여 새로운 감성모델을 만들었다. 또 Agarap(2020)은 감성 분류 분석에 LSTM(long-short term memory)을 가진 RNN을 사용하여 그래디언트가 사라지는 문제를 해결하였으며 이를 추천과 비추천의 감성 분류에 적용시킨 결과 0.88과 0.93의 F-1 score를 얻었다.

감성 분석 결과를 활용하여 다른 변수를 예측한 연구로 윤소영, 윤성대(2020)의 연구에서는 상품 리뷰를 분석하고 이를 가중치로 사용하여 협업 필터링에 적용하였다. 또 온라인 리뷰에서 매출을 예측한 Chong et al.(2016)의 연구에서는 온라인 리뷰의 양과 평점, 온라인 홍보전략 (무료 배송과 할인)과 더불어 리뷰 감성이 매출에 미치는 영향을 MLP(multilayer perception)로 분석하였다. 그 결과 온라인 리뷰, 온라인 홍보전략, 온라인 감성이 제품 매출에 미치는 영향은 각각 달랐으며 온라인 리뷰 수는 감성이나 할인과 상호작용 하였다.

이상에서 본 바와 같이 구매 리뷰 텍스트에서 감성 분석은 주요 키워드를 추출하거나 감성 예측 알고리즘을 구현하는 연구가 주를 이루었다. 각 연구에서 감성 분석은 구매 리뷰에서 유용한 정보를 얻는데 적합함을 알 수 있다.

## (2) 감성 예측 모델

구매 리뷰를 위한 텍스트 마이닝은 크게 두 부분으로 나뉘는데 하나는 임베딩에 의해 텍스트를 연산 가능한 벡터로 변환시키는 것과 처리한 말뭉치를 인공신경망 분석에 의해 분류하고 예측하는 모델을 만드는 것이다. 아래에서 본 연구에서 사용한 임베딩방법과 알고리즘에 대해 살펴보겠다(Aurelien, 2019).

### 1) 임베딩 기법

자연어 처리에서 인공신경망분석을 위한 임베딩 기법으로 CV(Count

Vectorization), TF-IDF, Word2Vec이 있다.

#### ■ Count Vectorization

Count Vectorization은 텍스트에 등장하는 단어의 빈도수를 파악해 카운트 벡터를 계산하고 단어 주머니(Bag of Words)를 만드는 방법이다.

#### ■ TF-IDF

TF-IDF는 문서 내 단어의 빈도(TF)와 여러 문서 내 단어 빈도(IDF)를 동시에 고려하여 문서 내 단어의 중요도를 나타내는 방법이다. TF-IDF는 단어 주머니에서 주제와 상관없는 단어가 등장하는 빈도가 커지는 것을 보완하기 위한 방법이다.

$$TF-IDF = w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

$tf_{i,j}$  = j 문서에서 I 단어의 출현빈도

N = 문서의 수

$df_i$  = 단어*i*가 발생하는 문서의 수

#### ■ Word2Vec

Word2Vec은 인공신경망 기반의 워드 임베딩 학습 모델로 비슷한 문맥을 가진 단어들이 공간에서 서로 가까이 분포하도록 한다. Word2Vec에는 CBOW(continuous bag-of-words)와 스킵그램(skip-gram) 방법이 있는데 CBOW는 주변 단어들을 이용하여 단어를 예측하는 방법이며 skip-gram은 현재 주어진 단어를 이용하여 주변 단어를 예측하는 방법이다.

## 2) 인공신경망 알고리즘

감성예측 모델에 사용할 수 있는 알고리즘으로 머신러닝에 SVC, logistic regression, random forest classifier, bagging classifier, multinomial NB 등이 있으며, 딥러닝 알고리즘을 CNN, RNN, LSTM 등이 있다. 각각을 살펴보면 다음과 같다(Aurelien, 2019).

#### ■ SVC

SVC는 분류에 사용하는 지도학습 머신러닝 모델로 각 학습 데이터 포인트가 두 클래스 사이의 결정 경계를 대표하는데 얼마나 중요한지를 학습한다. 클래스 사이의 경계에 높은 데이터 포인트인 서포트 벡터 (support vector)를 사용하여 결정 경계를 정의하고, 분류되지 않은 점을 해당 결정 경계와 비교해서 분류하는 방법이다.

■ logistic regression

로지스틱 회귀분석은 예측변수가 이항변수인 경우 연속 반응변수와의 관계를 근사화할 때 사용된다. 로지스틱 회귀분석은 선형회귀에서 예측변수가 연속변수가 아니라 이항 데이터인 경우 적용되지 못한다는 단점을 보완하여 예측변수가 0과 1사이에 있을 때 S자형 비선형 로지스틱 회귀선에 의해 분석하는 모델이다.

■ random forest classifier

랜덤포레스트는 분류분석에 사용되는 앙상블 학습 방법의 일종으로, 훈련 과정에서 구성된 다수의 결정 트리로부터 분류 또는 평균 예측치를 출력한다. 이때 예측 성능을 향상시키기 위해 상관관계가 없는 대규모 트리 컬렉션을 구축한다.

■ bagging classifier

배깅은 샘플을 부트스트랩으로 여러번 뽑아 각 모델을 학습시키고 결과물을 집계하는 방법이다. 먼저 데이터로부터 복원 랜덤 샘플링인 부트스트랩을 하고 부트스트랩한 데이터를 모델로 학습시킨 다음 결과를 집계하여 최종 결과 값을 구한다.

■ multinomial NB

나이브 베이스 모델은 데이터가 적을 때 간단하면서도 효율적으로 학습할 수 있다. 나이브 베이스 모델에서는 주어진 데이터가 특정 클래스에 속하는지를 확률을 통해서 예측하는데 텍스트 데이터와 같이 희소한 고차원인 경우 높은 정확도와 속도를 제공한다.

■ CNN(Convolutional Neural Network)

CNN은 일반적인 신경망 분석에서 이미지나 영상과 같은 데이터를 처

리할 때 발생하는 문제점을 보완한 방법이다. CNN은 convolution, pooling, flattening의 과정으로 구성되며 완전히 연결된 신경망을 학습시켜서 평면화시키는 과정에서 이미지의 공간 정보를 유지한 상태로 학습하는 알고리즘이다.

#### ■ RNN(Recurrent Neural Network)

순환신경망(RNN)은 입력과 출력을 연관된 연속 데이터인 시퀀스(Sequence)단위로 처리하는 모델로 자연어 처리나 순서가 있는 데이터 처리에 적합하다. RNN은 텍스트 분석에서 CNN이 구문 전체와 의미론적 관계가 복잡한 문장에 대해서 성능이 좋지 못함을 극복할 수 있다. RNN은 짧은 시퀀스를 처리할 때 유리하며 데이터의 시퀀스가 길어지면 역전파 알고리즘의 학습에 어려움이 있고 기울기 소실의 문제가 발생한다. 기울기 소실에 따라 과거 정보가 현재의 학습에 영향을 미치지 못하는데 이러한 문제점을 해결하기 위해 LSTM, GRU 등의 방법을 사용한다(김정미, 이주홍, 2017).

#### ■ bi-directional LSTM

LSTM(Long Short-Term Memory)은 RNN의 한 종류로 긴 기간의 시계열 데이터의 학습에 효율적이다. bi-directional LSTM(양방향 LSTM)이란 두 개의 LSTM으로 구성된 시퀀스 처리 모델로 하나는 입력을 순방향으로 하고 하나는 역방향으로 하여 네트워크에서 사용할 수 있는 정보의 양을 효과적으로 증가시키는 방법이다.

### 3) 임베딩 방법과 인공신경망 알고리즘

텍스트 마이닝에서 여러 임베딩 방법과 알고리즘을 결합하여 사용할 수 있다. Ali et al.(2019)는 SNS분석에서 단어 임베딩 기법으로 Word2Vec skip-gram과 퍼지온톨로지를 결합하여 벡터를 생성하고 Bi-LSTM 알고리즘을 사용하여 감성 분류 모델을 분석하였다. 연구에서 단어 임베딩에서 stringToWordVector와 TF-IDF를 결합한 방법, Doc2Vec, Glove2Vec, Word2Vec과 퍼지 온톨로지의 결합 방법을 비교한 결과 Word2Vec과 퍼지 온톨로지의 결합이 가장 정확도(accuracy)가 높게

나타났다. 또 인공신경망 분석 알고리즘으로 Bi-LSTM이 SVM, CNN, RNN보다 정확도가 높았다. 결론적으로 워드임베딩과 인공신경망 분석 알고리즘의 결합으로 Word2Vec과 퍼지온톨로지는 Bi-LSTM으로 분석했을 때 가장 정확도(accuracy)가 높았다.

박재영(2021)의 연구에서는 텍스트 분석에 있어서 인공신경망이나 트랜스포머 네트워크를 기반으로 한 BERT의 등장 이후 임베딩 모델이 크게 발전하였으나 모델이 대형화될수록 학습에 필요한 컴퓨팅 파워와 시간이 늘어나 보편화되기 힘든 면을 지적하며 TF-IDF 임베딩 벡터에 Word2Vec 유사도에 따른 가중치를 부여한 결합 벡터를 생성하여 문장-문서 비교를 통해 성능을 향상하였다.

## 제3절 구매 리뷰 분류

제3절에서는 구매 리뷰의 분류를 위한 유형화 연구를 살펴보고, 리뷰 유형을 사후적으로 해석하기 위한 방법으로 토픽분석을 살펴보겠다.

### (1) 구매 리뷰 유형화

온라인 구매 리뷰를 유형화한 연구에서는 내용분석과 같은 질적연구로 유형화하거나 양적 분석인 경우 주로 계층적 군집분석이나 k-평균 군집화를 사용하여 군집을 추출하였다. 또는 연결망 분석을 이용하기도 하였다(Kar et al., 2019; Kortiaty, et al., 2019; McKinney & Shim, 2016; 최영현, 이규혜, 2020; 허지욱, 2018). 온라인 쇼핑물의 구매 리뷰는 그 성격상 빅데이터인 경우가 많아 인공지능망 분석을 사용한 군집분석으로 비지도학습이 많이 이루어졌다. 군집분석은 비지도학습으로 군집 수를 연구자의 판단에 의해 결정해야 하고 분류분석이 재현될 때 분석의 정확도를 확인하기 어렵다는 단점이 있다. 특히 온라인 리뷰는 그 성격상 개인의 견해가 표현되기 때문에 유동성이 커서 이러한 유동성을 개념화하는 것이 필요하다는 점을 감안할 때 온라인 리뷰의 유형화 연구는 아직도 해결해야 할 부분이 많은 도전적인 연구 주제다(Kawarf & Istanbuluoglu, 2019).

온라인 구매 리뷰의 유형화 연구 사례는 다음과 같다. 질적 연구로 McKinney and Shim(2016)은 패션 렌탈 경험에 대한 리뷰 600개를 하나씩 코딩하여 내용분석으로 3개의 유형을 분리하여 온라인 리뷰를 구조화하는데 기여하였다. 최영현과 이규혜(2020)은 의미 네트워크 분석을 통해 만족 요소와 불만족 요소를 군집화 하였다. 계층적 군집분석을 사용하여 온라인 리뷰 내용을 유형화한 연구로 Kar et al.(2019)의 연구에서는 에어비앤비에 대한 온라인 리뷰를 2011년부터 2015년까지 5년간의 영국 데이터 196,666개를 계절 및 시간의 흐름에 따라 좋은 품질에 대한 생각이 바뀌었는지 분석하였다. 이때 분석은 top 10 키워드에 대해 텍스트 마이

닝으로 TF-IDF를 구하고 계층적 군집분석, 시계열 분석 등으로 온라인 리뷰의 시간적 변화를 분석하였다. 그 결과 시간이 지나도 공통적으로 좋은 품질의 요건이 되는 요소가 있는 반면 특정 년도와 시즌에 독특하게 나타나는 속성도 발견되었다.

온라인 리뷰는 빅데이터를 얻을 수 있으므로 온라인 리뷰의 유형화에 인공지능망 모델을 적용한 연구에서는 주로 Word2Vec을 이용하여 단어 간 유사도를 통해 임베딩하고 이를 K-평균 군집화등의 방법으로 유형화하였다. 김장경과 박은혜(2017)는 이러닝 수강 후기를, 허지욱(2018)은 네이버 블로그의 글을 Word2Vec을 이용하여 단어 간 유사도를 구하고 임베딩하여 KNN과 같은 클러스터링 알고리즘으로 군집화하였다. 강형석과 양장훈(2019)은 Word2Vec으로 단어를 임베딩하고 군집화한 후 군집의 특성 및 t-SNE 분포로 단어간 관계 표현 능력을 정성적으로 분석하였다. 권성범 (2020)은 아마존 리뷰를 군집분석한 후 3개의 상품 유형 즉 하드웨어 제품, 서비스 상품, 하드웨어와 클라우드 서비스가 융합된 형태의 상품 유형별 리뷰가 서로 다른 군집에 속함을 확인하였다.

## (2) 토픽분석

구매 리뷰의 유형화 이후 각 유형별 내용을 분석하기 위해 토픽모델을 사용할 수 있다. 토픽모델은 문서에서 추상적인 주제를 발견하기 위해 문서 내 단어의 분포를 분석하여 주제를 파악하는 기법으로 LSA(Latent Semantic Analysis)와 LDA(Latent Dirichlet Allocation)를 많이 사용한다 (이지현 외, 2020).

Kortiat, et al.(2019)은 항공 승객의 온라인 리뷰를 구조적 토픽모델(structural topic models)와 LDA로 분석하여 서비스 품질의 다차원적인 측면과 만족의 주요 요인을 밝혔다. 이지현 외(2020)은 온라인 리뷰 분석을 통한 상품 평가 기준 추출하기 위해 LDA 및 k-최근접 이웃 접근법을 활용하였다.

### (3) 유형화 모델

리뷰 텍스트의 유형화를 위해 사용할 수 있는 군집화 모델은 정답의 라벨이 없고 데이터의 분포가 비슷한 것끼리 분류하는 비지도 학습 모델이다. 본 연구에서는 구매 리뷰의 유형화를 위해 K-평균군집화 알고리즘을 사용하였다. K-평균 군집화 알고리즘은 2차원 이상의 벡터공간으로 표현된 데이터를 K개의 집단으로 군집화하는 알고리즘으로 임의의 K개의 중심점(centroid)를 정하고 코사인 유사도나 유클리디언 거리를 구해 각 데이터와 중심점 간의 거리를 최소가 되게 하는 집단을 구하는 방법이다. K-평균군집화의 단점은 초기 중심점에 의해 군집화 결과가 달라지며, 군집의 모양이 중심점을 중심으로 한 구형으로 제한되며, 중심점과 노이즈와의 거리가 멀 경우, 노이즈에 의해 잘못된 centroid가 학습된다는 것이다.

K-평균군집화는 비지도 학습으로 군집의 개수를 결정할 때 보통 elbow method로 중심과의 거리의 제곱합이 군집의 수에 따라 더 이상 크게 감소하지 않는 점을 선택한다. 군집화 결과를 평가하는 방법으로 사용되는 실루엣 계수는 -1에서 1사이의 값을 가지는데 평균값이 0에 가까울수록 근처의 군집과 가깝고, 1에 가까울수록 근처의 군집과 더 멀리 떨어져 있어 군집화가 잘되었다고 할 수 있다. 실루엣 계수의 평균값 1에 가까울수록 좋으며 실루엣 계수의 군집 간 편차가 작은 것이 좋다.

$$s(i) = \frac{(b(i) - a(i))}{(\max(a(i), b(i)))} \quad (2)$$

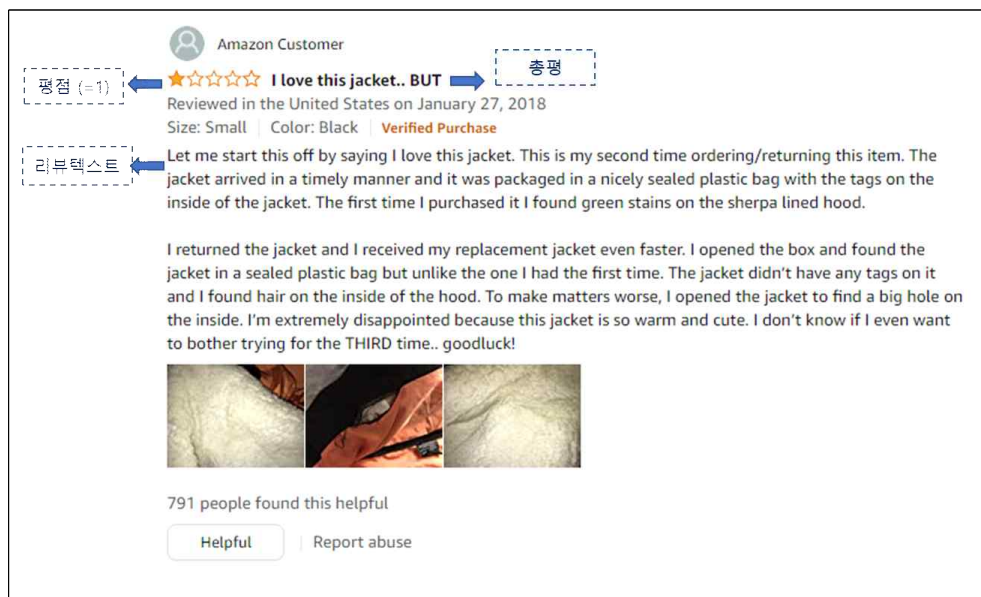


### III 장 연구결과 분류 및 예측 모델 개발

#### 제1절 데이터 수집 및 전처리

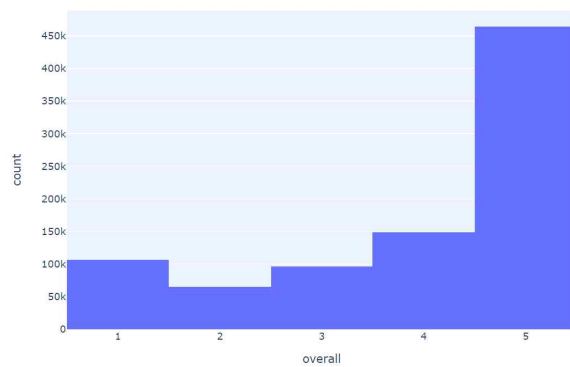
##### (1) 데이터 수집

본 연구에 사용한 데이터는 Amazon review dataset 2018에 공개된 Clothing\_Shoes\_and\_Jewelry\_5 데이터 셋을 사용하였다. 데이터는 12개의 열로 구성되었으며 평점, 구매확정, 리뷰시간, 리뷰ID, asin code, 리뷰어 이름, 리뷰텍스트, 총평, 표준시간, 좋아요, 상품 스타일, 상품이미지에 대한 정보를 포함한다. 본 연구에서는 평점과 리뷰 텍스트 데이터 및 총평을 분석에 사용하였다. 데이터 분석환경은 Windows 11 운영체제와 파이썬 기반의 아나콘다의 Jupyter Notebook를 활용하였다. 아마존 리뷰 데이터셋에서 리뷰 텍스트와 총평, 평점의 위치는 <그림 2>와 같다.



<그림 2> 아마존 리뷰 예시

데이터의 수는 모두 883,636개이었으며 그 중 리뷰텍스트, 총평에서의 결측치 1,741개(0.20%) 제외하고 881,895개의 데이터를 사용하였다. 데이터에서 평점은 <그림 3>에서 보듯이 1점에서 5점까지 분포하며 5점이 가장 많아 464,262개(52.6%)이었으며 4점이 149,103개(16.9%), 3점이 96,945개(11.0%), 2점이 64,669개(7.3%), 1점이 106,916개(12.1%)로 분포하였다.



<그림 3> 데이터셋의 평점별 분포

## (2) 데이터 전처리 (pre-processing)

### 1) 데이터 선택 및 분류

본 연구에서는 평점을 타겟으로 하여 분석하고자 한다. 평점의 분포에서 가장 긍정적인 반응인 5점에 52.7%가 분포하는 반면 부정적인 반응인 평점인 1점과 2점을 합해도 171,798개(19.4%)에 지나지 않았다. 따라서 평점 1점과 2점을 합쳐서 부정으로 라벨링하고 5점을 긍정으로 라벨링하였다.

긍정과 부정 평점의 분포를 고르게 하기 위해 'shuffle'로 데이터를 섞고 텍스트 리뷰의 결측치 1,741개 (0.20%)를 제외한 총 881,895개의 데이터를 분석에 사용했다. 이 중 평점 3점, 4점의 중간 점수를 제외하고, 1점과 2점인 데이터 171,585개 (19.5%)와 5점인 데이터 464,262개(52.6%)를 합해 모두 635,847개(72.1%)의 데이터를 사용하였다. 한편 긍정과 부정 리뷰의 수를 균등하게 하여 모델의 성능을 높이기 위해 긍정과 부정 데이터를 1:1 로 투입하여 분석하였다. 또한 사용하는 데이터

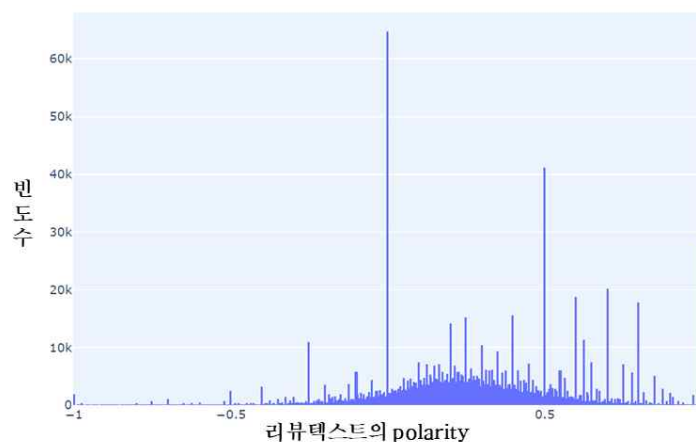
셋에서 훈련데이터와 시험데이터의 비율을 8:2로 하여 분류하였다.

## 2) 데이터 전처리

데이터셋에서 리뷰텍스트에 대해 데이터 임베딩을 위한 전처리를 하였다. 먼저 모든 문자를 소문자로 바꾸고 구두점과 숫자를 제거하였다. 그다음 토큰화(tokenization)하고 불용어(stopwords)를 제거하였다. 또 표제어 추출(lemmatization)을 하였고 길이가 2 이하인 단어는 제거하고 나서 마지막으로 토큰을 문장으로 다시 결합하여 데이터 파일을 만들었다. 또한 텍스트 데이터에서 특성을 추출(feature extraction)하기 위해 토큰화하고 임베딩으로는 bag-of-words 기법을 사용하는 CV(Countverization)와 TF-IDF(Term Frequency-Inverse Document Frequency), Word2Vec을 사용하였다.

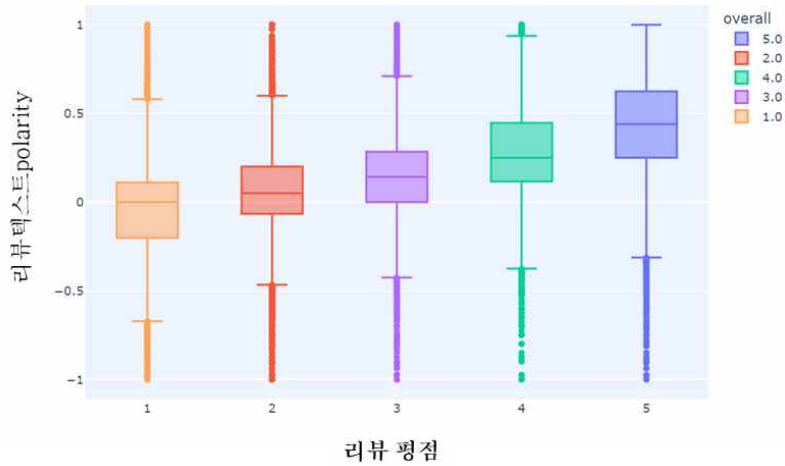
## 3) TextBlob을 사용한 감성분석

텍스트에 대해 전처리한 데이터를 TextBlob를 사용하여 리뷰 텍스트에 대해 polarity를 구하였다. 리뷰 텍스트와 총평의 polarity 분포는 <그림 4>, <그림 5>와 같다. 리뷰 텍스트의 polarity와 평점과의 관계는 평점이 높을수록 polarity 평균이 높아 긍정적인 것으로 나타났다 <그림 4>. 리뷰 텍스트의 평점별 polarity 분포에



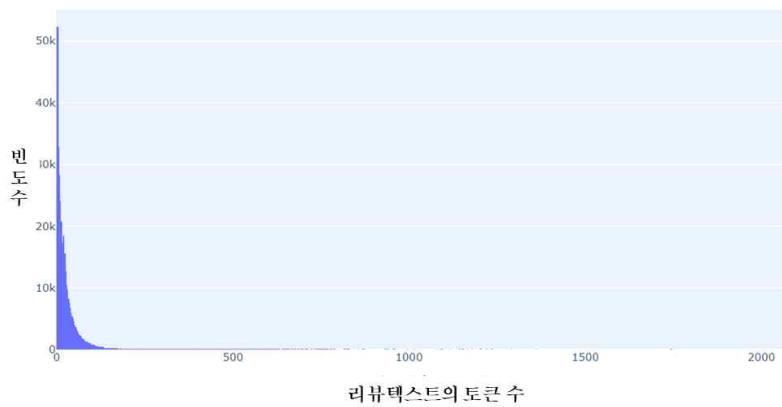
<그림 4> 리뷰 텍스트와 총평의 polarity 분포

서는 1점에서 5점으로 가면서 polarity가 증가하였다 <그림 5>.



<그림 5> 평점 별 리뷰 텍스트의 polarity 분포

리뷰 텍스트를 토큰화 했을때 토큰의 길이 분포를 <그림 6>과 같이 살펴본 결과, 길이의 분포가 왼쪽에 극단적으로 치우쳤으며 오른쪽 꼬리가 급하게 떨어짐을 알 수 있다. 이를 통해 리뷰를 짧게 쓴 사람들의 분포가 길게 쓴 사람보다 높았으며 리뷰 길이의 분산이 크지 않음을 알 수 있다.

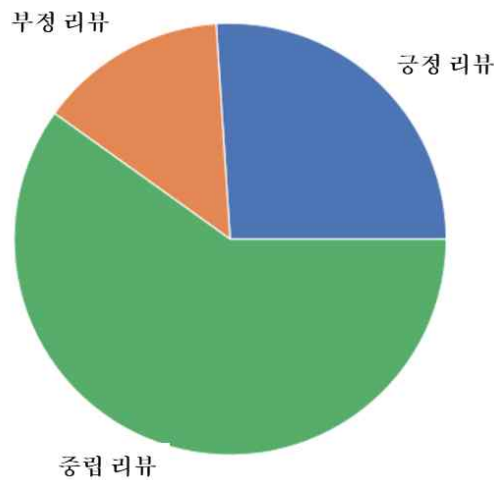


<그림 6> 리뷰 텍스트의 토큰 수 분포

리뷰 텍스트의 polarity를 긍정과 부정으로 나누어 보기 위해 polarity가 1일 때 긍정, 0.5일 때 중립, 0이하일 때 부정으로 하였다. 그 결과 <표 1>과 <그림 7>과 같이 전체 데이터에서 리뷰 텍스트의 polarity의 분포는 중립에 절반 이상(55.6%)의 데이터가 분포 하였다. 긍정 대 부정 반응의 비율은 긍정 리뷰가 부정 리뷰보다 1.8배 많았다.

<표 1> 리뷰 텍스트의 polarity 분포

polarity	긍정 (p=1)	중립 (p=0.5)	부정 (p<0)
리뷰 텍스트 분포 (%)	24.2%	55.6%	13.1%



<그림 7> 텍스트 리뷰 polarity의 긍정, 부정, 중립 분포

텍스트 리뷰의 polarity의 긍정, 중립, 부정에서 각각 샘플 문장을 10개씩 뽑아 보았다. 그 결과는 <표 2>, <표 3>, <표 4>과 같다. 긍정보다 부정 리뷰에서 문장이 긴 경우를 더 많이 발견할 수 있었고 이로부터 부정적인 리뷰를 쓰는 사람이 글을 더 길게 쓰는 경향이 있다고 추측할 수 있다. 이는 조용희 외(2018)의 연구에서 영화에 대해 불만족스러운 경우 즉 평점이 낮을 때 만족스러운 경우보다 영화 리뷰의 길이가 길었던 것과 일치하는 결과이다.

<표 2> 긍정 텍스트 리뷰 샘플

긍정 텍스트 리뷰 샘플 (10개)

excellent

Beautiful!

Great price!

The best shop from I can get any color any time I want,love the delivery time and material,

used this jacket while traveling to Dublin. Great purchase!

I love it!!!! Thank you!!! It looks just like the dress and it fits perfectly!!!

Excellent quality

Perfect for my newborn baby, born in the summer and who loves spending time outdoors.

The best price for this model! Delivered as a set, not just a frame. Excellent.

Beautiful! Great buy!!

<표 3> 중립 텍스트 리뷰 샘플

중립 텍스트리뷰 샘플(10개)

Love it.

This top fits more like a Woman's Medium.

great thanks

Love it

This is a cheap garment I don't like it at all!

Love the outfit

Nice fit

Comfy and I get lots of compliments :)

It's cute and comfortable!

Love it

#### <표 4> 부정 텍스트 리뷰 샘플

##### 부정 텍스트리뷰 샘플(10개)

HIDEOUS!! SO SO UGLY!!! Very boxy cut, no shape to it. I looks like a mechanics jumpsuit from a Nascar pit stop. Not denim/chambray fabric AT all; it's like a cheap blue cotton that's not even soft. Save your money

did not fit well.

Too tight, and the belt hoops were at hip level instead of normal waist level.

Maybe others bought from a different vendor but mine is awful. The material is not cotton like or mixed as stated and too small. I am 5'4" and it's too short also I fit in size 6 pants and the leg opening is too small. The worst is the pocket placement, might as well not have any pockets. The pockets are a good 8" from the waistband. Not a good purchase, run away.

These seem to do the trick in the pool, but I'm not super psyched about how small they run or how tight they seem around the waste and legs of my son. They leave deep red marks after every pool session and he outgrows them super fast because they do run so small.

Weird fabric, cape was smaller than in the pic, no belt

I bought this as a gift for my sister who's obsessed with butterflies and she loves it

It runs small. Is very thin. I'll likely never wear it because the quality is not what I expected.

Way too small. Must have been mismarked. Immediately sent it back.

Way to big and crazy long.

#### 4) Top-ngram을 이용한 단어 빈도

Top-ngram으로 CountVectorizer로 가장 출연 빈도가 높은 단어 순으로 단어 수 한 개, 두 개, 세 개인 경우를 각각 20개씩 구하여 다음과 같이 살펴보았다.

##### ■ 한 개 단어 Top-ngram

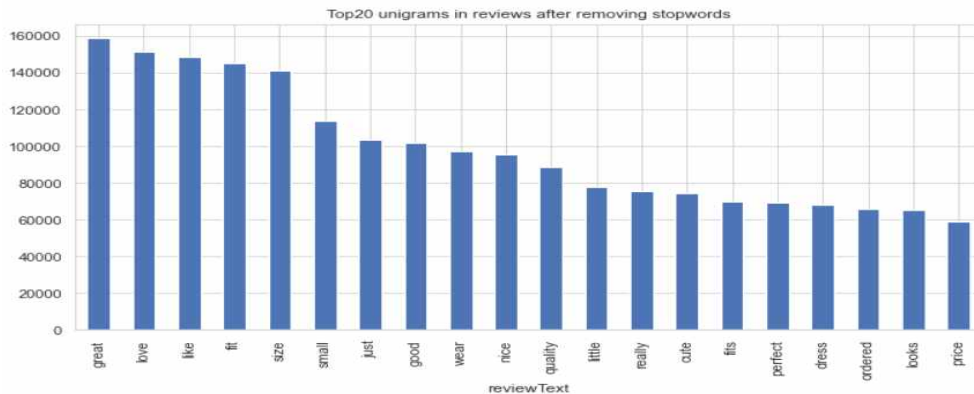
리뷰 텍스트에서 Count Vectorizer로 단어 수를 계산하여 빈도가 가장 높은 단어 20개를 구한 결과 <그림 8>과 같이 great, love, like, fit, size, small, just, good, wear, nice, quality, little, really, cute, fits, perfect, dress, ordered, looks, price가 나왔다. great, nice, good 등 좋다는 표현 외에도 size, small, fit 등의 단어가 많이 나타나 옷의 맞춤새와 사이즈가 구매 후 만족에 중요함을 알 수 있었고 특히 큰 사이즈보다 옷이 작을 경우 문제가 됨을 리뷰 내용으로부터 알 수 있었다.

■ 두 개 단어 Top-2gram

리뷰 텍스트에서 CountVectorizer로 단어수를 계산하여 같이 등장하는 두 개의 단어쌍 중 빈도가 가장 높은 단어 20개를 구한 결과 <그림 9>와 같이 good quality, year old, like picture, looks like, super cute, great quality, fit perfectly, looks great, way small, look like, fits perfectly, ordered size, great price, really, like, true size, fit great, fits great, great product, love love, just like 등이 가장 빈도가 높은 단어로 나타났다.

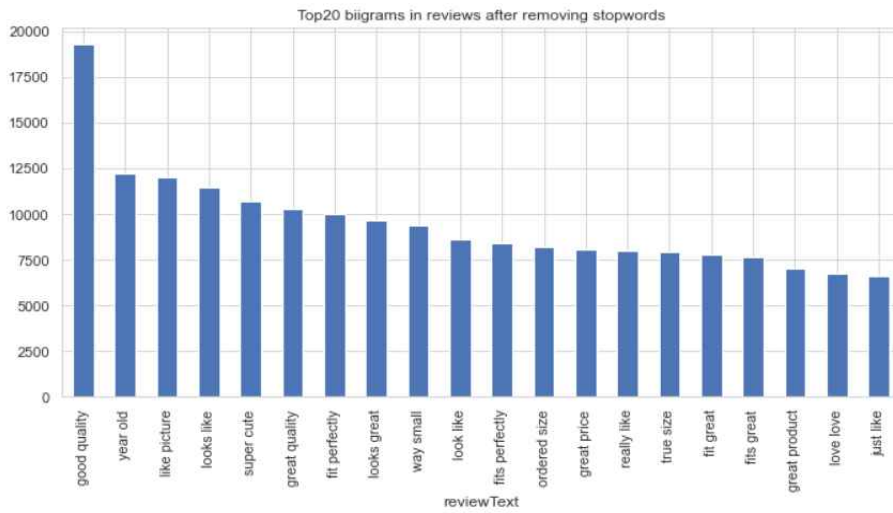
■ 세 개 단어 Top-3gram

텍스트리뷰와 총평에서 CountVectorizer로 단어수를 계산하여 같이 등장하는 세 개의 단어 쌍 중 빈도가 가장 높은 단어 20개를 구한 결과 텍스트리뷰에서는 “love love love”, “just like picture”, “looks just like”, “honest unbiased review”, “look like picture”, “exchange honest review”, “exchange honest unbiased”, “looks like picture”, “discount exchange honest”, “year old daughter”, “exactly like picture”, “don waste money”, “received product discount”, “looks exactly like”, “doesn look like”, “does look like”, “fit just right”, “fit true size”, “product discount exchange”, “year old son” 으로 나왔다. 내용은 전체적인 만족, 사진과 제품의 일치에 대한 내용이 많았다. 그 외에도 총평에서는 가성비에 대해 많이 언급했고 사이즈에 대해 언급했다<그림 10>.

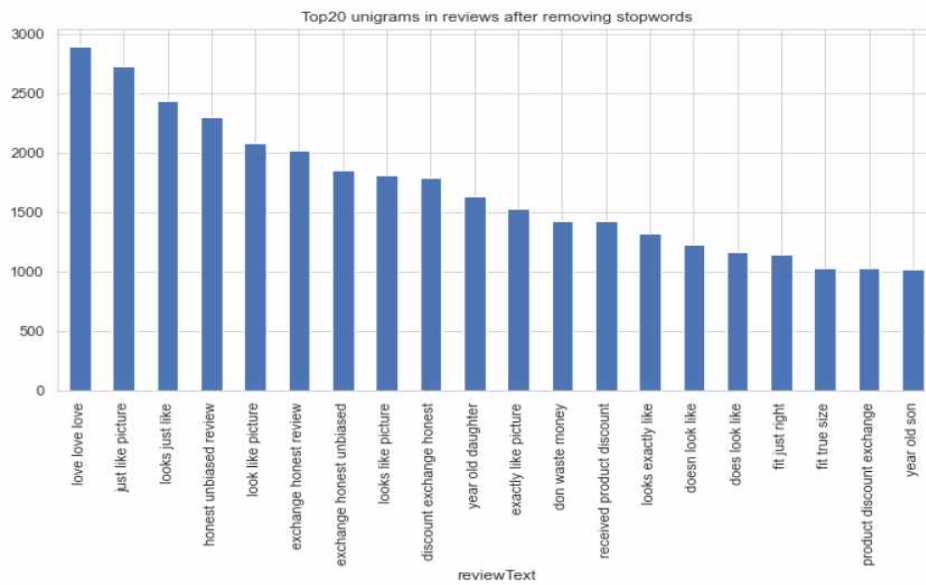


<그림 8> 리뷰 텍스트의 한 개 단어 top-ngram





<그림 9> 리뷰 텍스트의 두 개 단어 top-ngram



<그림 10> 리뷰 텍스트의 세 개 단어 top-ngram

## 제2절 알고리즘 구축

### (1) 데이터 임베딩

텍스트 데이터를 분석하기 위해 텍스트를 숫자로 바꾸는 임베딩 과정이 필요하다. 각 알고리즘에 적합한 다양한 임베딩 방법을 사용하여 리뷰 텍스트 분석에 가장 효율적인 방법을 찾고자 하였다. 토큰의 개수를 세서 수치화 하는 CountVectorization이나 문서내 빈도수를 고려한 TF-IDF를 사용하여 머신러닝 알고리즘으로 리뷰 텍스트의 감성을 분석하였다. 또한 원핫인코딩에 의해 sparse한 데이터를 연속형 변수로 바꾸는 방법으로 Word2Vec을 사용하여 결과를 비교하였다. Word2Vec의 경우 skip gram을 사용하였고 gensim model을 적합시켰다. Bi-LSTM 알고리즘의 학습을 위해 pad to sequence와 text to sequence로 임베딩 하였다. 각 임베딩 방법은 학습 알고리즘과 결합하여 사용하였다.

### (2) 학습 알고리즘

리뷰 텍스트의 유형화하기 위해 k-평균 군집화, DBSCAN 알고리즘을 구축하여 성능을 테스트하고 비교하였다. 한편 리뷰 텍스트로부터 감성을 예측하는 모델을 만들기 위해 SVC, logistic regression, random forest classifier, bagging classifier, multinomial NB 등의 머신러닝 알고리즘과 bi-directional LSTM 알고리즘을 사용하여 테스트하고 비교하였다. 임베딩 방법으로 CV, Tf-idf, word2vec, doc2vec, pad-sequence와 알고리즘의 조합은 <표 5>와 같다. 분류모델은 k-평균 군집화와 DBSCAN으로 분석하였으며 이때 word2vec을 사용하였다. 또 군집화 후 토픽을 LDA, LSA를 사용하여 분석하였으며 이때 리뷰 텍스트는 Count Vectorization으로 하였다. 예측모델에서 SVC, logistic regression, random forest classifier, bagging classifier은 Count Vectorization, Tf-Idf, word2vec으로 임베딩하여 비교하였고, CNN, RNN 모델을 word2vec으로 임베딩하여 비교하였다. 또 Bi-directional LSTM 모델을 pad-sequence로 임베딩하여 테스트 하였다.

<표 5> 데이터 임베딩과 학습 알고리즘 테스트

알고리즘 \ 임베딩		CV	Tf-Idf	word2vec	post-sequence
분류모델	k-평균 군집화			O	
	DBSCAN			O	
	LSA	O			
	LDA	O			
예측모델	SVC	O	O	O	
	logistic regression	O	O	O	
	random forest classifier	O	O	O	
	bagging classifier	O	O	O	
	multinomial NB	O	O		
	CNN			O	O
	RNN			O	
	Bidirectional LSTM				O

## 제3절 예측 시뮬레이션

### (1) 감성 예측 모델

#### 1) 데이터 전처리

리뷰 평점을 예측하는 모델을 구성하였는데 예측데이터는 리뷰텍스트와 총평을 비교하였다. 평점값은 원래 데이터에서는 1부터 5까지의 값을 가지나 평점 5에 52.6%로 반이 넘는 데이터가 분포하고, 4점에는 149,103개(16.9%), 3점이 96,945개(11.0%), 2점이 64,669개(7.3%), 1점이 106,916개(12.1%)로 분포한다. 따라서 데이터의 긍부정을 예측하기 위해 평점 5를 긍정으로 보고 평점 1점과 평점 2점에 해당하는 데이터를 부정으로 보았다. 그럼에도 불구하고 평점 1점과 평점 2점 데이터의 합이 총 19.4%에 지나지 않아 52.6%를 차지하는 평점 5와는 데이터 양에 차이가 크다. 그래서 본 연구에서는 평점 1점과 평점 2점을 부정으로 코딩하고 평점 5점을 긍정으로 코딩하되 데이터를 shuffle 로 섞고 위에서부터 같은 크기를 선택하였다. 이러한 방법으로 긍정과 부정 데이터를 같은 수로 사용하여 모델을 예측하였다.

#### 2) 데이터 임베딩

데이터셋에서 리뷰 텍스트에 대해 데이터 임베딩을 위한 전처리를 하였다. 먼저 모든 문자를 소문자로 바꾸고 구두점과 숫자를 제거하였다. 그 다음 토큰화(tokenization)하고 불용어(stopwords)를 제거하였다. 또 표제어 추출(lemmatization)을 하였고 길이가 2 이하인 단어는 제거하고 나서 마지막으로 토큰을 문장으로 다시 결합하여 데이터 파일을 만들었다. 특성 추출을 위해 토큰화하고 CV(Countvectorization), Tfidf, doc2vec을 사용하여 리뷰 텍스트를 임베딩 하였다.

#### 3) 감성예측 모델

리뷰 텍스트로부터 평점을 예측하는 모델을 구축하기 위해 학습데이터와 테스트데이터의 비율은 8:2로 하였다. 온라인 리뷰로부터 평점을 예측하는 모델을 구축하고 성능을 평가하기 위해 CV, Tfidf, Word2vec으로 임베딩한 값을 SVC, lo-

gistic regression, random forest classifier, bagging classifier, multinomial NB의 알고리즘을 사용하여 모델링하였다. 또 텍스트 분석에 효율적인 bi-directional LSTM을 모델링하였다.

## (2) 성능평가

### 1) 머신러닝 모델

리뷰 텍스트를 Count Vectorization(CV), tf-idf, word2vec으로 임베딩하고 머신러닝 알고리즘으로 SVC, logistic regression, random forest, bagging, multinomialNB 이용하여 모델링하고 각 모델의 정확도를 비교하였다. 그 결과 <표 6>과 같이 SVC와 logistic regression에서는 Tf-Idf 임베딩 방법에서 정확도가 0.924로 가장 높았다. 한편 random forest classifier와 multinomial NB 모델에서는 CV 임베딩 했을 때 정확도가 0.910과 0.93으로 가장 높았다. 또한 bagging classifier에서는 word2vec으로 임베딩했을 때 정확도가 0.890으로 가장 높았다. 전체적으로 보면 머신러닝 알고리즘 모델에서는 CV와 TfIdf 임베딩 방법이 word2vec보다 우수한 것으로 보였으며 각 알고리즘에 따라 적합한 임베딩 방법에 차이가 있음을 알 수 있다.

<표 6> 모델 정확도

알고리즘 \ 임베딩		CV	Tf-Idf	word2vec	post-sequence
머신러닝	SVC	0.909	0.924	0.886	.
	logistic regression	0.921	0.924	0.886	.
	random forest classifier	0.910	0.909	0.901	.
	bagging classifier	0.875	0.874	0.890	.
	multinomial NB	0.93	0.92	.	.
딥러닝	CNN	.	.	0.696	.
	Bidirectional LSTM	.	.	.	0.94

### 2) CNN 모델 (Word2Vec 임베딩)

긍정, 부정 평점에 따라 각각 300,000개의 데이터를 선택해 CNN모델기를 사용하

여 분석하였다. 이때 <표 7>과 같이 gensim Word2Vec으로 임베딩하고 모델링에서 Convolutional layer (activation: relu), Max pooling, Dense (10, activation=relu), Dense (1, activation=sigmoid)를 사용하고 optimizer를 'adam', loss 함수를 'binary\_crossentropy'를 사용하여 ephcos=10, batch size=10으로 모델링한 결과 정확도가 0.696이 나왔다. 모델의 파라미터는 <표 8>과 같다. 훈련데이터와 테스트데이터의 epoch에 따른 정확도와 손실은 <그림 11>과 같다.

<표 7> CNN 모델

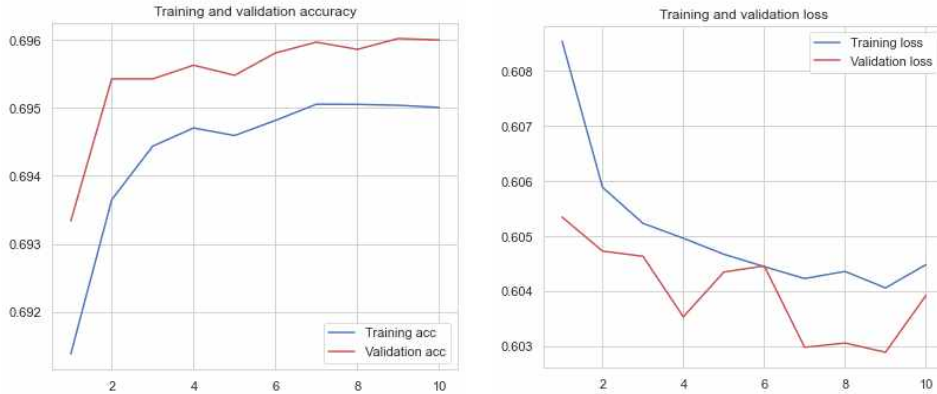
특성 추출	embedding	gensim Word2Vec
데이터분할	훈련데이터:학습데이터	8:2
모델링	Convolution	(128,5, activation=:relu)
	Pooling	GlobalMaxPooling1D
	Dense	(128, activation: relu)
	Dense	activation : sigmoid
	optimizer	adam
	loss	binary_cross_entropy

<표 8> CNN 모델 파라미터

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 377327, 100)	37732700
conv1d (Conv1D)	(None, 377323, 128)	64128
global_max_pooling1d (Global	(None, 128)	0
dense (Dense)	(None, 10)	1290
dense_1 (Dense)	(None, 1)	11

```
Total params: 37,798,129
Trainable params: 37,798,129
Non-trainable params: 0
```



<그림 11> 에포크에 따른 데이터 정확도

### 3) LSTM (post-sequence 임베딩)

데이터는 리뷰평점이 3,4로 중간인 것과 결측치를 제외하고 긍정과 부정을 각각 70,000개 씩 총 70,000개가 사용되었다. 모델학습과 테스트를 위해 데이터를 훈련데이터와 평가 데이터를 8:2로 나누어 112,000개의 훈련데이터와 28,000개의 평가데이터를 분석에 사용하였다. 리뷰 텍스트를 불용어 처리한 후 pad-sequence로 토큰화하였다. 제일 긴 리뷰는 624, 제일 짧은 리뷰는 1개의 단어로 이루어졌으며, 사용된 총 단어수는 22,290개이다. 이때 padding을 사용하여 데이터 길이를 200으로 균등하게 길이를 맞추었다. 또 긍정 부정의 인덱스를 text-to-sequence로 토큰화 하였다. 모델은 bi-directional sequential LSTM을 사용하였으며 하이퍼 파라미터는 <표 9>와 같이 단어수는 5000, 임베딩 차원은 64, LSTM 셀의 규모는 100, 기사의 길이는 200, 학습용 데이터의 비율은 0.8, 반복 학습수(epochs)으로 하였다. 모델은 Dencs=6, activation은 softmax, optimizer는 adam, loss 함수는 sparse-categorical crossentropy를 사용하였다. 모델 요약은 <표 10>과 같다. 모델의 정확도는 0.94 이고 loss는 0.093으로 나왔다. 혼동행렬은  $\begin{bmatrix} 13246 & 755 \\ 850 & 13149 \end{bmatrix}$ 으로 나왔다.

<표 9> LSTM 모델

특성 추출	리뷰 텍스트	pad-sequence
	감성(긍정, 부정) index	text-to-sequence
데이터분할	훈련데이터:학습데이터	8:2
임베딩	sequential	64차원
모델링	model	bidirectional LSTM (units = 100)
	embedding	input_dim = 5000(MY VOCAB) 임베딩 차원 64 (dropout rate=0.5)
	dense	3차원
	activation	softmax
	optimizer	adam
	loss	sparse-categorical_crossentropy
	epochs	10

<표 10> LSTM 모델 파라미터

RNN 요약

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, None, 64)	320000
dropout (Dropout)	(None, None, 64)	0
bidirectional (Bidirectional)	(None, 200)	132000
dense (Dense)	(None, 6)	1206

Total params: 453,206

Trainable params: 453,206



<표 11> 리뷰 감성 예측 샘플

리뷰어 (날짜)	유형	내용	평점	예측
Kelpie (May 20, 2018)	리뷰	<p>These socks are well made, and fit well - no lumpy toes. They are just the right weight to comfort my tender feet, but not thick like boot socks. The only drawback is the wide elastic inside the cuff. It invariably causes deep indentations in my ankles, cutting off circulation and causing my feet to swell. I have medical issues that causes a tendency toward this problem. I am sure most people would be happy with them, The colors are fun and cheerful.</p> <p>They go through the laundry looking like new. The colors have stayed bright even after several washings. If you wear shoes bigger than 8.5 you might find them too snug, or not quite able to fit as securely over the heel as you would prefer.</p>	4	긍정
	총평	Comfortable, durable, a very good buy.		긍정
E Poole (Jan. 16, 2018)	리뷰	<p>I received these socks and am quiet pleased with them, at least the pairs that made it through the first washing. After receiving them and before wearing them, I put them in the wash after which a hole appeared in one of the socks at the heal corner where they come together with the foot and the ankle. I tried emailing the seller and received no response in regard to a replacement sock for the one that fell apart. Overall synopsis: cute and warm if they make it through the first wash. Expect 20% to fall apart and no response from the seller to rectify the situation.</p>	2	부정
	총평	Nice socks if they make it through the first wash		부정
Beth (Oct. 10, 2017)	리뷰	<p>I love these socks! They're thick, super comfy, soft and fit my size ten feet well. They wash up nicely with regular detergent and do not shrink or pill. The label recommends washing in warm water but I used hot and everything was fine. The colors are vibrant and great for fall and winter.</p>	5	긍정
	총평	Attractive and Comfortable		긍정
Mrs. Kristine White (Oct. 26, 2018)	리뷰	<p>JUST TERRIBLE! These socks are horrible! Scratchy and tight and I could be more disappointed! What in the world? I am so upset! And NO returns? How I regret this purchase</p>	1	부정
	총평	Waste of money! So scratchy and fit very tight. Gross.		부정

LSTM 모델의 성능을 테스트 하기 위해 아마존 쇼핑에서 패션제품 (삭스)에 대한 리뷰를 랜덤으로 가져와서 성능을 테스트해보았다(사이트 URL: [https://www.amazon.com/YZKKE-Womens-Vintage-Winter-Multicolor/dp/B075D9G69J/ref=sr\\_1\\_2?qid=1641795012&s=fashion-womens-intl-ship&sr=1-2&th=1](https://www.amazon.com/YZKKE-Womens-Vintage-Winter-Multicolor/dp/B075D9G69J/ref=sr_1_2?qid=1641795012&s=fashion-womens-intl-ship&sr=1-2&th=1)). 데이터에 포함되지 않은 임의의 아마존 사이트 댓글 4명의 댓글에서 리뷰텍스트와 총평을 각각 가져와 8개를 테스트해본 결과 <표 11>와 같이 8개 모두 올바르게 예측하였다.

### (3) 감성별 토픽분석

온라인 패션물 상품구매 리뷰를 평점에 의해 긍정(평점이 5점 중 5점) 과 부정(평점이 5점 중 1점 또는 2점)으로 나누어 토픽 분석을 하였다. 긍정평점의 리뷰 15,000개, 부정평점의 리뷰 15,000개를 분석에 사용하였다. 온라인 리뷰를 평점으로 긍정리뷰와 부정 리뷰로 나눈후, 각각에 대해 토픽분석을 하였다. 온라인 리뷰 데이터의 임베딩은 Count Vectorization을 사용하였다.

#### 1) 긍정 감성 리뷰 토픽

온라인 리뷰 중 긍정적인 평점에 대한 리뷰를 LSA와 LDA 방식으로 토픽 분석했다.

##### ① LSA (Latend Semantic Analysis)

LSA로 구한 8개의 토픽은 <표 12>와 같고 각 토픽에서 주요 단어로 나온 top 3 word는 <그림 12>와 같다. 가장 많이 나온 토픽은 “great fit size”로 압도적으로 많은 비중을 차지했다. 토픽의 수와 t-SNE는 <그림 14>와 같다.

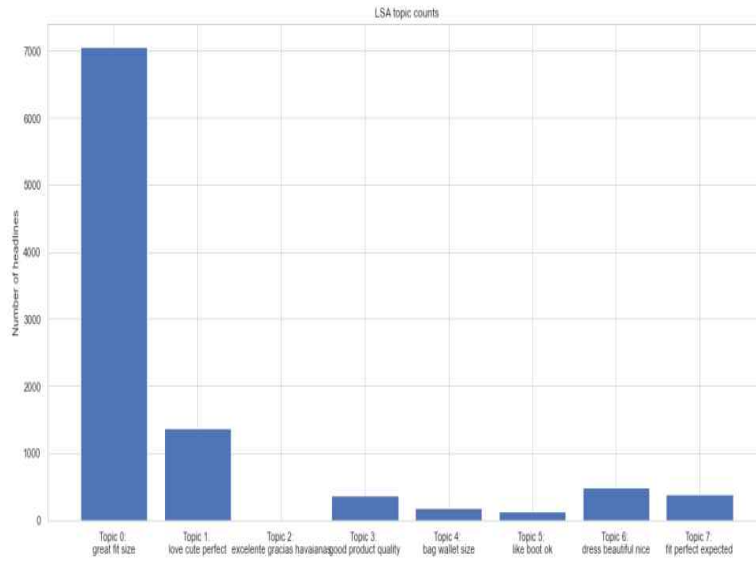
##### ② LDA (Latent Dirichilet Allocation): great love fit perfect fits size cute comfortable just nice)

LDA로 구한 8개의 토픽은 <표 12>와 같고 각 토픽에서 주요 단어로 나온 top 3 word는 <그림 13>과 같다. 가장 빈도가 높게 나온 토픽은 “great love fit”이었고

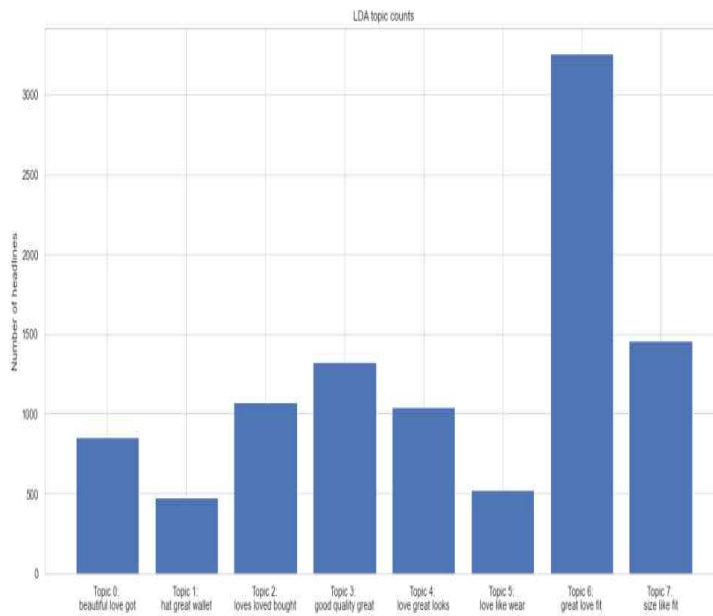
그 다음 “size like fit”이 높게 나왔고, “good quality great”, “loves loved bought”, “love great looks”, “beautiful love got”등이 나왔다. 토픽의 수와 t-SNE는 <그림 15>와 같다.

<표 12> 긍정 리뷰의 토픽

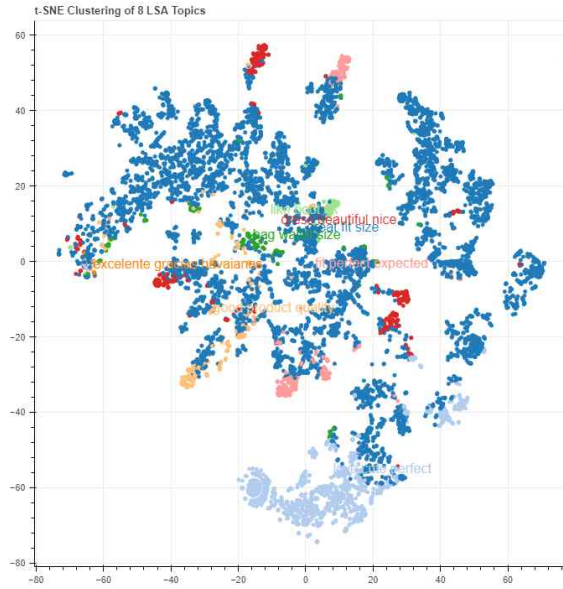
topic	LSA	LDA
1	great fit size love just like quality wear perfect nice	beautiful love got great necklace compliments just recommend absolutely pretty
2	love cute perfect beautiful color comfortable fits wear fit absolutely	hat great wallet love purse bag like sun perfect head
3	excelent gracias havaianas wrong shhhh experimental explaining exposing exposed expose	loves loved bought great daughter gift ring old perfect son
4	good product quality price nice like wallet really excellent looking	loves loved bought great daughter gift ring old perfect son
5	bag wallet size perfect purse great leather pocket small pockets	love great looks like quality wear beautiful nice ring perfect
6	like boot ok wear feel sole boots comfortable lot wearing	love like wear great pants warm comfortable perfect just long
7	dress beautiful nice fits perfect great quality daughter like size	great love fit perfect fits size cute comfortable just nice
8	fit perfect expected perfectly comfortable good cute loved dress like	size like fit love dress just little small wear ordered



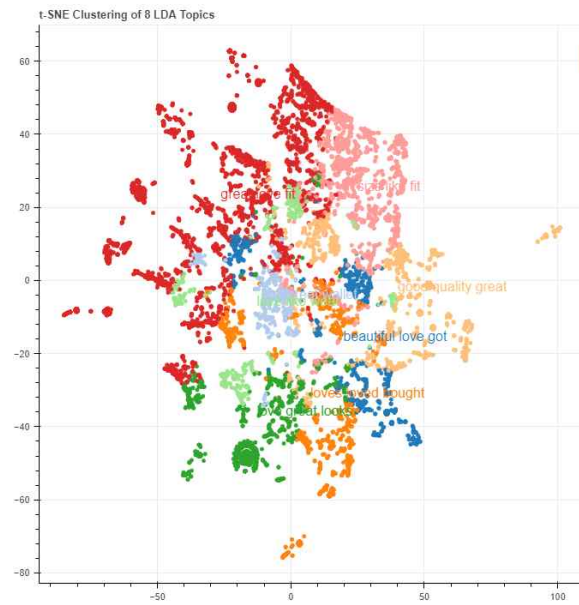
<그림 12> LSA에 의한 긍정 리뷰 토픽



<그림 13> LDA에 의한 긍정 리뷰 토픽



<그림 14> LSA에 의한 긍정리뷰의 t-SNE



<그림 15> LDA에 의한 긍정리뷰의 t-SNE

## 2) 부정 감성 리뷰 토픽

부정적인 평점에 대한 리뷰를 LSA와 LDA 방식으로 토픽 분석했다.

### ① LSA (Latend Semantic Analysis)

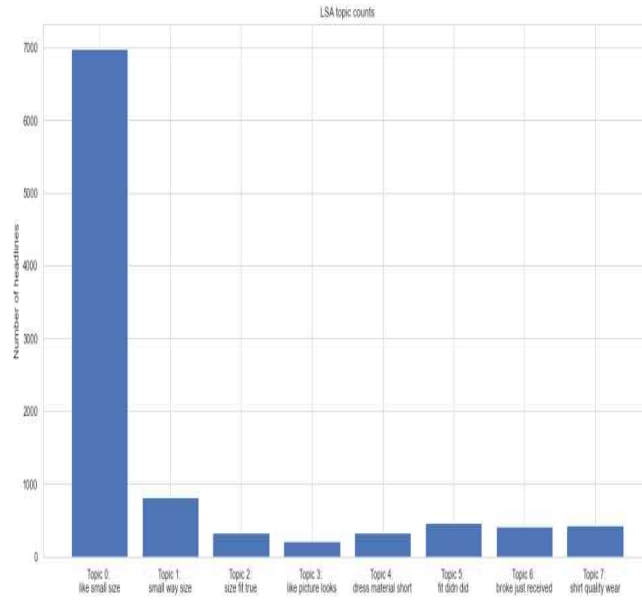
LSA로 구한 8개의 토픽은 <표 13>과 같고 각 토픽에서 주요 단어로 나온 top 3 word는 <그림 16>과 같다. 가장 빈도가 높은 토픽은 “like small size”가 압도적으로 높았고 사이즈에 대한 것이 대부분 이었다. 토픽의 수와 t-SNE는 <그림 18>과 같다.

② LDA (Latent Dirichilet Allocation): great love fit perfect fits size cute comfortable just nice)

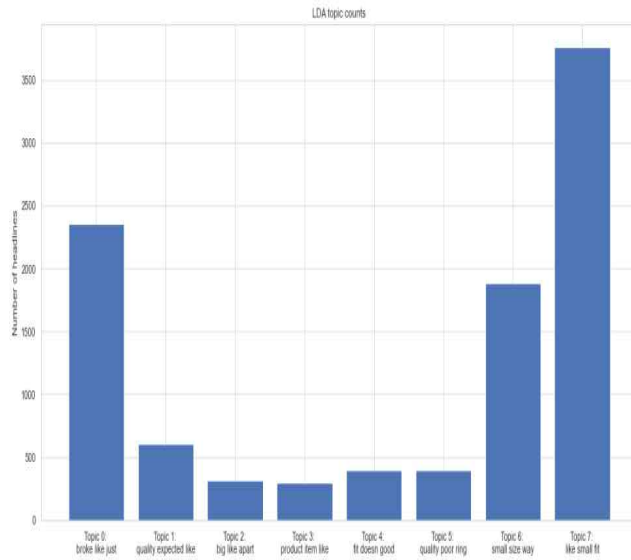
LDA로 구한 8개의 토픽은 <표 13>과 같고 각 토픽에서 주요 단어로 나온 top 3 word는 <그림 17>과 같다. 가장 빈도수가 높은 토픽은 “like small fit”이었고 그 다음으로 “small size way”, “broke like just”로 사이즈나 핏에 관한 불만이 중요한 것으로 나타났고 특히 작은 사이즈가 문제되는 것을 볼 수 있다. 그 외에 “quality”에 관한 토픽이 있었다. 토픽의 수와 t-SNE는 <그림 19>와 같다.

<표 13> 부정 리뷰의 토픽

topic	LSA	LDA
1	like small size cheap fit quality material wear way ordered	broke like just quality product cheap time good wear don
2	small way size ordered runs large sizes run medium cute	quality expected like color hat disappointed small cute cheaply cheap
3	size fit true ordered smaller chart big order disappointed large	big like apart really just small came product wearing smell
4	like picture looks look does doesn photo fits xl looked	product item like terrible small size recommend quality purchased sweater
5	dress material short like cheap small size fit look picture	fit doesn good like does small cut size look long
6	fit didn did does doesn like way small wear big	quality poor ring fell wore worth wearing like small cheap
7	broke just received day time fell wear got week broken	small size way ordered large fit wear like medium sizes
8	shirt quality wear poor small like large ordered picture material	like small fit dress size picture material cheap looks shirt



<그림 16> LSA에 의한 부정 리뷰 토픽



<그림 17> LDA에 의한 부정 리뷰 토픽

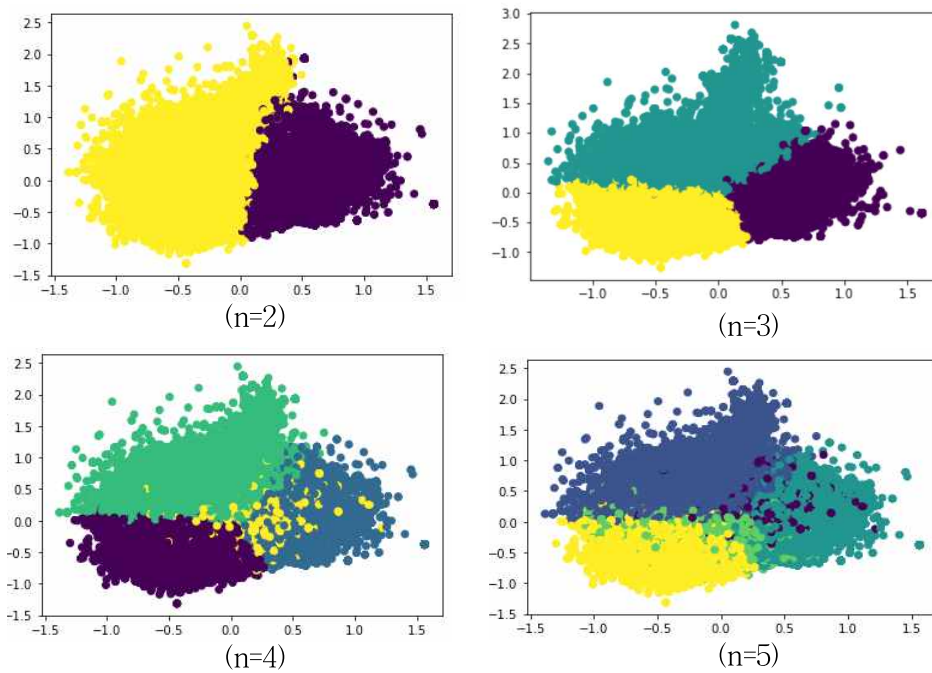




## 제4절 유형 분류

### (1) 리뷰 유형 분류

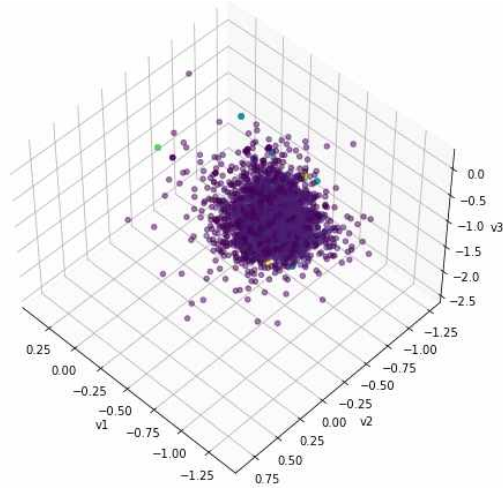
리뷰텍스트를 분류하기 위해 word2vec으로 20차원으로 임베딩한 데이터를 사용하여 k-평균 군집화를 했다. <그림 20>과 같이 PCA를 이용한 scatter plot으로 살펴보면 군집 수에 따른 데이터 분포에서 군집 수가 2와 3일 때 잘 분리되었고, 총 설명 분산을 높이면서 집단 간의 구별이 잘 되는 3개로 분류 집단 수로 선택하였다.



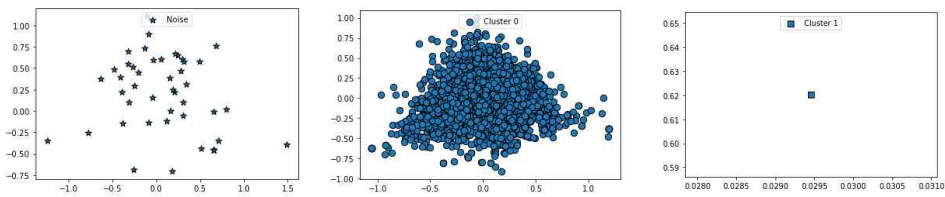
<그림 20> 군집 수에 따른 k-평균 군집화 결과

군집화 알고리즘으로 군집의 수를 정하지 않고 데이터의 모양에 따라 군집을 찾는 방법으로 DBSCAN을 사용하였는데 그 결과 <그림 21>과 같이 데이터가 한군데로

뭉쳐 있었으며 <그림 22>와 같이 구분할 수 있는 경계면을 찾지 못하여 집단이 잘 분리되지 않았다.



<그림 21> 데이터의 3-D 점 표현



<그림 22> DBSCAN 결과

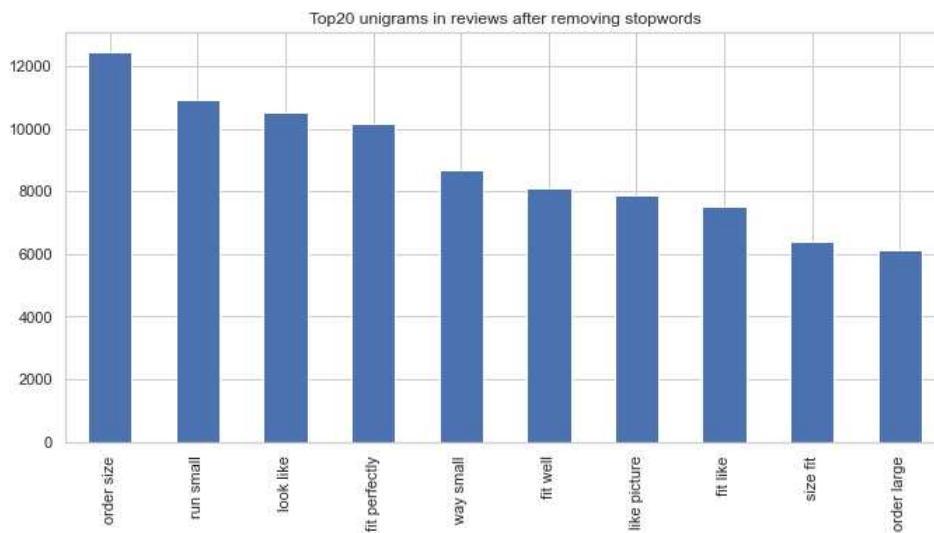
리뷰 텍스트를 3개의 집단으로 clustering 했을 때 각각 한 단어, 두 단어, 세 단어에 대해 n-gram을 10개씩 구해서 각 집단의 특성을 살펴보았다. 그 결과 두 단어를 추출 했을 때 가장 해석 가능한 군집의 특성을 볼 수 있었다. 리뷰 텍스트의 단어를 3개의 집단으로 군집화한 결과 각 집단의 top-2gram을 10개씩 뽑아본 결과는 <표 14>와 같다.

집단 1은 사이즈와 맞춤새에 대한 내용이 많아 'size & fitting'으로 명명했다. 집단 2는 품질과 가격에 대한 것이 많았는데 싼가격, 비싼가격, 우수한 품질, 가격대비 품질 등 가격과 품질 등에 대한 긍정적인 댓글과 부정적인 댓글이 다 있었으며

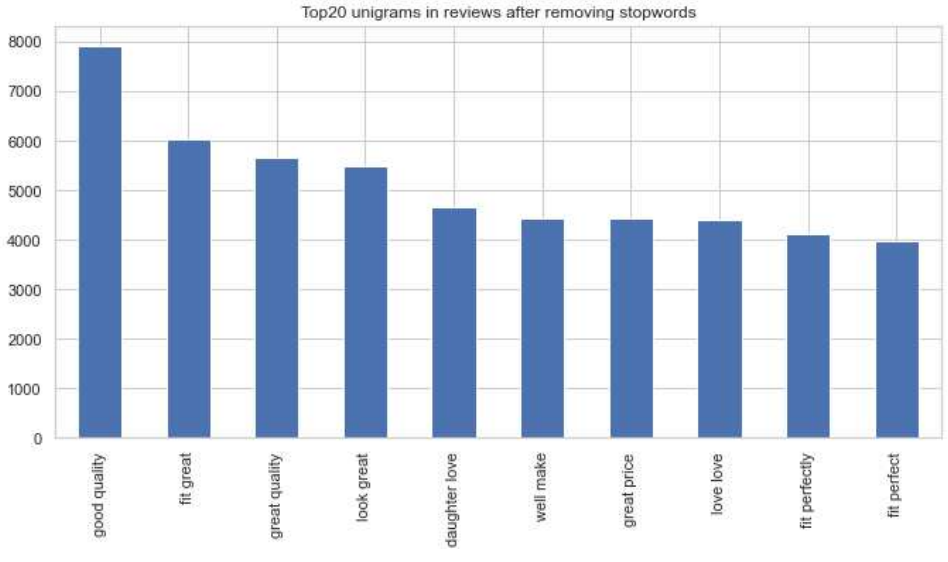
‘quality & price’로 명명했다. 집단 3은 look well, look great, look good, like picture와 같은 외형에 대한 댓글이 많아 ‘appearance’로 명명했다. 군집화한 3개 집단에서 n-gram으로 뽑은 대표 단어는 <그림 23>, <그림 24>, <그림 25>와 같다. 각 군집 내의 대표 단어들은 의미가 서로 연결되었으며, 그렇지 않은 몇몇 단어들은 k-평균 군집화 실행 시의 노이즈로 볼 수 있다.

<표 14> 리뷰 유형별 n-gram

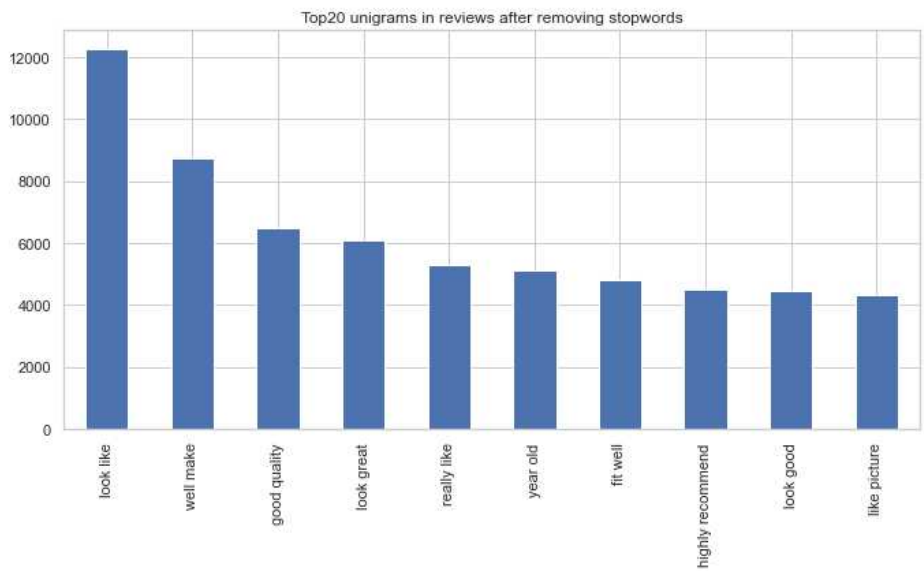
cluster	n	평점	2-gram
집단 1 Size & Fit	284,245	3.33	order size, run small, look like, fit perfectly, way small, fit well, like picture, fit like, size fit, order large
집단 2 Quality & Price	242,934	4.83	good quality, fit great, great quality, look great, daughter love, weel make, great price, love love, fit perfectly, fit perfect
집단 3 Appearance	353,308	3.73	look like, well make, good quality, look great, really like, year old, fit well, highly recommend, look good, like picture



<그림 23> ‘Size & Fit’ 그룹의 n-gram



<그림 24> 'Quality & Price' 그룹의 n-gram



<그림 25> 'Appearance' 그룹의 n-gram

## (2) 성능 평가

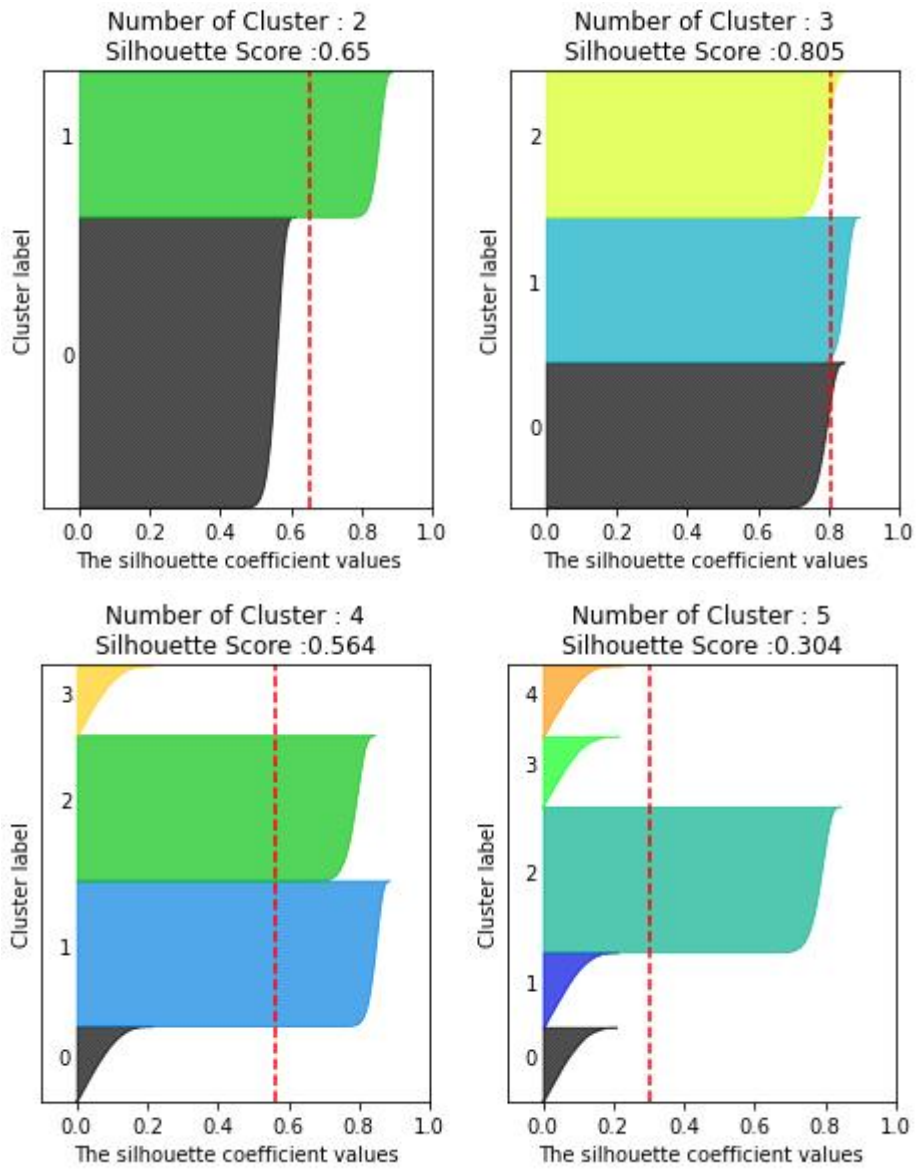
온라인 리뷰 분류의 유효성을 평가하기 위해 실루엣 스코어, ANOVA, pairwise tukeyhsd 사후검정을 구하고, 분류 집단 별로 평점 예측 모델의 성능을 평가하였다. 또 총평이 리뷰 분류 결과의 집단을 예측하는 정도를 LSTM모델로 측정하였다.

### 1) 실루엣 점수

군집화 분석의 성능을 평가하기 위해 실루엣 점수를 사용했다. 리뷰 텍스트로 k-평균 군집화한 결과 군집 수에 따른 실루엣 점수의 변화를 <그림 26>과 같이 살펴보았다. 군집 수에 따른 실루엣 점수는 군집이 3일 때 모든 집단이 평균을 넘고 분산이 치우치지 않음을 알 수 있다. 또 3개의 집단으로 유형화했을 때 실루엣 스코어가 0.805로 가장 높았으며 집단 간의 크기도 균형을 이루었다.

### 2) ANOVA 검정

리뷰 텍스트를 Clustering한 결과 나온 3개의 유형별로 평점에 차이가 있는지를 ANOVA 분석을 하였다. 각 유형별 평균은 <표 15>와 같이 집단1('Size & Fit')의 평균은 3.33, 집단2('Quality & Price')의 평균은 4.83, 집단3('Appearance')의 평균은 3.73이다. 이들 평균 점수의 유형 간 차이를 보기 위해 ANOVA 분석 결과 <표 16>과 같이 F값이 94166.24(p=0.0)으로 3개의 평점 값의 평균에 유의한 차이가 있는 것으로 나타났다. 리뷰 유형별 집단 간 평점 평균 값의 차이가 유의한지 검정하기 위해 사후검정으로 pairwise tukeyhsd를 하였다. 그 결과 <표 17>과 같이 집단 간 차이가 유의하게 나왔다(alpha=0.05).



<그림 26> 실루엣 점수

<표 15> 집단 별 평점 평균

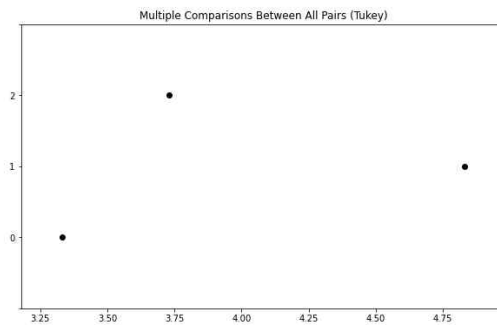
군집	이름	N (데이터수)	평점 평균(5점만점)
0	'Size & Fit'	284.245	3.332
1	'Quality & Price'	242.934	4.831
2	'Appearance'	353.308	3.729

<표 16> ANOVA

	df	sum_sq	mean_sq	F	PR(>F)
C(군집)	2.0	3.12e+05	156162.33	94166.24	0.0
Residual	880484.0	1.46e+06	1.66	NaN	NaN

<표 17> 평균의 다중비교: Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	1.499	0.001	1.490	1.507	True
0	2	0.397	0.001	0.390	0.405	True
1	2	-1.101	0.001	-1.109	-1.093	True



<그림 27> 집단 평균

### 3) 분류 유형별 평점 예측 성능

분류 유형의 적정성을 평가하기 위해 각 분류 집단 별로 평점 예측 성능을 평가하였다. 연구문제 2-2 에서 개발한 LSTM알고리즘을 활용하여 긍부정 평점 예측 모델의 성능을 평가하였다. 그 결과 전체 데이터에 LSTM 모델을 분석한 경우 정확도가 0.94인데 비해 각 집단에 대해 LSTM 모델을 분석한 결과 세 집단에서 모두 정확도가 0.93으로 나왔다. 한편 각 세분 집단 별 모델 평가를 위해 아마존 패션제품 리뷰 8개를 랜덤으로 가져와서 성능을 테스트한 결과 세 집단에서 모두 8개 샘플 다 긍부정을 정확하게 예측하였다. 이상과 같이 유형분류의 안정성을 보여주었다.

### 4) 총평에 대한 유형 분류 성능

본 연구에서 사용한 AMAZON FASHION 데이터셋에서 리뷰 텍스트 대신 총평을 이용하여 리뷰 텍스트로부터 나온 유형 분류의 재현성을 테스트하였다. 총평을 임베딩하여 LSTM 알고리즘으로 유형 예측 모델을 적합시킨 결과 0.79의 정확도를 얻었다.



## 제 IV 장 결론

### 제1절 요약 및 시사점

본 연구는 온라인 쇼핑몰의 패션상품 구매 리뷰의 빅데이터 분석을 통해 감성예측 모델을 만들고 리뷰의 내용을 유형화하여 의미론적으로 분석하고자 했다. 온라인 패션몰 상품 구매 평점을 바탕으로 감성예측 모델을 만들기 위해 CV, Tf-Idf, Word2Vec의 임베딩 방법과 SVC, logistic regression, random forest classifier, bagging classifier, multinomial NB, CNN, bi-directional sequential LSTM 알고리즘을 테스트하였다. 성능 평가 결과 bi-directional sequential LSTM 모델이 가장 성능이 우수한 것으로 나타났다. 패션상품 구매 리뷰를 의미론적으로 분석하기 위해 k-평균 군집화로 구매 리뷰를 세 개의 집단으로 분류하였다. 그 결과에 따라 패션상품 구매 리뷰를 “Size & Fit”, “Quality & Price”, “Appearance”의 세 개 유형으로 분류하였다. 각 군집 별 평점 평균은 “Quality & Price” 그룹이 가장 높았고, “Size & Fit”와 “Appearance” 그룹은 낮게 나왔다. 이는 온라인에서 패션상품을 구매할 때 물건을 직접 보거나 착용하지 못하고 구매해야 하는 한계점을 말해준다고 할 수 있겠다.

패션상품 구매 리뷰 분류의 유효성을 평가하기 위해 실루엣 스코어, ANOVA, pariwise tukeyhsd 사후검정을 구하고, 분류 집단 별로 평점 예측 모델의 성능을 평가하였다. 또 총평이 리뷰 분류 결과의 집단을 예측하는 정도를 LSTM모델로 측정하였다. 결론적으로 패션상품 구매 리뷰는 bi-directional LSTM 알고리즘으로 감성을 효율적으로 예측할 수 있으며, 패션상품 구매 리뷰는 그 특성에 따라 3가지 유형으로 분류할 수 있다.

본 연구는 온라인 패션상품 구매 리뷰의 분류 및 예측의 틀을 제시함으로써 향후 연구에 필요한 기초자료를 제공했다. 온라인 상품구매 리뷰에 다양한 내용이 섞여 있는 경우가 많아 자세히 읽어 보기 전에는 대표적인

내용을 추출하거나 긍·부정 감성을 파악하기 힘들며 때때로 거짓 평점으로 인해 리뷰와 평점이 일치하지 않는 경우가 있다. 이같이 온라인 상품 구매 리뷰 데이터가 빅데이터임에도 불구하고 유용하게 사용되기 힘들고 고려할 때 본 연구와 같이 bi-directional sequence LSTM 알고리즘을 사용하여 구매 리뷰로부터 높은 정확도로 감성 예측이 가능함을 보여 준 것은 상품 기획 및 마케팅에서 실무적 의의가 크다. 패션 상품 리뷰의 감성 예측 모델을 통한 실시간 모니터링으로 해당 상품에 대한 긍·부정 감성을 조기에 예측하는 것이 가능하며 이에 따라 효율적인 수요예측 및 재고 관리가 가능하다. 또 평점과 댓글의 감성이 일치하지 않을 때 평점만으로 정확히 파악할 수 없는 소비자 만족도를 댓글을 통해 직접 예측함으로써 효율적인 대응이 가능하다.

본 연구에서 상품 리뷰를 유형화함으로 패션 상품 구매와 관련된 소비자 관심 영역 및 만족 불만족의 결정요인을 파악할 수 있다. 패션상품 리뷰의 유형화 결과 “Size & Fit”, “Quality & Price”, “Appearance”의 세 집단으로 분류하였다. 이로부터 온라인 패션상품 구매에서 소비자가 구매 후 평가하는 기준으로 사이즈와 핏, 품질, 가격, 외형 등이 중요시 한다는 것을 알 수 있다. 소비자 구매의사결정 시 평가 기준에 대한 연구는 많으나 본 연구에서는 위와 같이 구매 후 평가 기준을 제시했다는 의미가 있다. 또한 온라인리뷰를 유형화한 선행연구에서 주로 군집분석을 많이 사용했는데 군집분석은 비지도학습으로 군집수를 연구자의 판단에 의해 결정해야 하고 분류분석이 재현될 때 분석의 정확도를 확인하기 어렵다는 단점이 있다. 특히 온라인 리뷰는 그 성격상 개인의 견해가 표현되기 때문에 유동성이 커서 이러한 유동성을 개념화하는 것이 필요하다는 점을 감안할 때 온라인 리뷰의 유형화 연구는 아직도 해결해야 할 부분이 많은 도전적인 연구 주제다(Kawarf & Istanbuluoglu, 2019). 본 연구에서는 온라인 리뷰 유형화의 어려움을 보완하기 위해 실루엣 스코어, ANOVA, 사후검정과 같은 통계수치를 구해 군집 수의 적정성을 검증하고자 하였고 토픽분석을 통해 유형화된 군집의 유동성을 개념화 설명하고자 하였다.

또 군집의 재현성을 평가하기 위해 리뷰 텍스트에서 분류된 내용을 총평에 적용시켜 보았다. 패션상품 영역에서의 온라인 리뷰 유형화를 위한 본 연구의 시도는 실무에서 유동성이 높은 패션 구매 리뷰를 조기에 파악하고 관리하는데 유용하게 쓰일 수 있을 것이다.

## 제2절 연구의 한계 및 제언

본 연구에서는 아마존 데이터셋을 이용하여 영어로 된 구매 리뷰를 분석하였는데 향후 한국어로 된 리뷰를 대상으로 본 연구 주제를 적용시켜 볼 필요가 있다. 본 연구의 한계점은 k-평균 군집화에 의해 패션상품 리뷰의 유형을 도출했음에도 불구하고, 개별 데이터의 유형을 분류하는 정답이 없기 때문에 분류의 효율성을 정확히 판단하기 힘들었다는 것이다. 향후 연구에서는 패션상품 리뷰의 유형을 직접 라벨링 하기 힘든 한계를 극복하여 유형할수 있는 프로토콜을 제시할 필요가 있겠다.

본 연구에서는 Word2Vec으로 임베딩하여 k-평균 군집화로 군집을 추출하였는데 품사별로 단어를 추출하여 군집화 한다던가, 주제별 쏠림을 균등화 하는 등 다양한 방법으로 보다 효율적으로 군집화할 수 있는 방안을 모색할 필요가 있겠다. 또 구매 리뷰와 더불어 총평을 비교 분석하여 각각이 소비자 반응 예측에 어떻게 다른 기능을 하는지에 대한 연구가 필요하며, 구매 리뷰를 상품 정보 및 소비자 정보와 연동시켜 연구하는 것이 필요할 것이다.

본 연구에서는 감성예측 모델 검증을 위해 다양한 임베딩 방법과 알고리즘을 사용하여 우수한 성능을 증명하였으나 분류모델의 성능을 우회적으로 평가하였다. 향후 연구에서는 여러 임베딩 방법을 합성하여 가중치를 주던지 전이학습 등의 방법을 사용하여 보다 높은 성능의 분류 모델을 개발하고 또 분류 모델의 성능을 평가할 수 있는 다양한 방안을 개발하는 것이 필요하겠다. 또 다양한 질적, 양적 연구방법을 보완하여 리뷰의 분류 시스템을 만들고 인공지능망분석을 통한 분석과 비교하여 패션 리뷰 분류의 프로토콜을 만드는 것이 필요하겠다.

## 참고문헌

- 강형석, 양장훈(2019). Word2Vec 모델로 학습된 단어 벡터의 의미 관계 분석, *정보과학논문지*, 46(10), 1088-1093.
- 김건영, 이창기(2016). “Convolutional Neural Network를 이용한 한국어 영화평 감성 분석”, *한국정보과학회 학술발표논문집*, 2016(6), 727-749.
- 김성범 (2020). “제품, 서비스, 융합제품서비스의 소비자 니즈 비교 분석: 아마존 온라인 리뷰를 중심으로”, *한국콘텐츠학회논문지*, 20(7), 316-300.
- 김장경, 박은혜(2017). “빅데이터 분석을 이용한 이러닝 수강 후기 분석,” *한국정보통신학회논문지*, 21(1), 423-428.
- 김정미, 이주홍(2017). Word2vec을 활용한 RNN기반의 문서 분류에 관한 연구, *한국지능시스템학회 논문지*, 27(6), 560-565.
- 김준겸, 최은솔, 윤수현, 이유빈, 김동환 (2021). 치킨 리뷰의 이면: 텍스트 마이닝을 통한 리뷰의 탐색적 분석을 중심으로. *한국콘텐츠학회 논문지*, 21(11), 30-40.
- 박재영(2021). 단어 간 유사도 기반 임베딩을 결합한 TF-IDF 알고리즘 개선에 관한 연구, 서울과학대학원대학교 석사학위논문.
- 봉원재, 방한별, 이지형(2016). 오토인코더를 활용한 유사도 기반 영화 추천 시스템, *한국지능시스템학회 학술발표 논문집*, 26(1), 61-62.
- 신승재, 나병후, 신동혁, 나영연(2017). 인공 신경망 기반 뉴스기사의 토픽 분석 및 문서 분류 동시 수행 모델 개발, *한국정보과학회 학술발표논문집*, 1876-1878.
- 윤소영, 윤성대 (2020). 상품 리뷰 감성분석을 이용한 아이템 기반 협업 필터링 추천 기법, *한국정보통신학회논문지*, 24(8), 970-977.
- 윤재연 (2019). 추가 정보를 고려한 상품 리뷰 요약 기법, 서울대학교 석

사학위논문.

- 이동엽, 조재춘, 임희석(2017). 워드 임베딩을 이용한 아마존 패션 상품 리뷰의 사용자 감성 분석, *한국융합학회논문지*, 8(4), 1-8.
- 이윤주, 소현정, 광기영, 안현철(2020). 텍스트 마이닝 기법을 활용한 웹툰 댓글 분석: 네이버 베스트 도전 웹툰을 중심으로, *한국컴퓨터정보학회 학술발표논문집*, 28(2), 219-222.
- 이정원, 박철(2020). 온라인 리뷰 특성이 영화 매출액에 미치는 영향: 한국과 미국의 비교, *경영학연구*, 49(6), 1575-1594.
- 이지현, 정상형, 김준호, 민은주, 여운영, 김종우 (2020). 온라인 리뷰 분석을 통한 상품 평가 기준 추출: LDA 및 k-최근접 이웃 접근법을 활용하여, *한국지능정보시스템학회*, 26(1), 97-117.
- 이호근, 광현 (2013). 온라인 소비자 리뷰의 효과에 영향을 미치는 요인에 대한 고찰. *정보화정책*, 20(3), 3-17.
- 조용희, 박이슬, 김혜진 (2018). 빅데이터를 활용한 영화 흥행에 따른 리뷰 길이 변화, *한국콘텐츠문화학회지*, 18(5), 367-375.
- 주보라, 황선진 (2016). 소셜커머스에서 부정적 리뷰 유형, 브랜드 명성, 기회회소성지각이 패션제품 선호도에 미치는 영향, *패션비즈니스*, 20(4), 207-225.
- 최영현, 이규혜 (2020). 겨울왕국2의 콜라보레이션 패션제품에 대한 소비자 리뷰 -의미 네트워크와 감성분석-, *복식문화연구*, 28(2), 265-284.
- 채희주, 박수현, 고은주 (2016). 서포터즈의 온라인 리뷰 유형에 따른 패션 브랜드의 온라인 인상형성과 구전효과에 대한 연구, *한국의류산업학회지*, 18(1), 15-26.
- 통계청(2021.10.5.). 2021년 8월 온라인쇼핑 동향, *온라인쇼핑동향조사*. <https://kostat.go.kr>, (retrieved on Dec.,10, 2022).
- 한기향(2021). 반팔 티셔츠의 온라인 리뷰 분석에 관한 융합적 분석 연구, *한국과학예술통합학회*, 39(4), 541-555.

- 허지욱(2018), Word2Vec을 이용한 한국어 단어 군집화 기법, *한국인터넷 방송통신학회 논문지*, 18(5), 25-30.
- Aurelien, G. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow: concepts, tools, and techniques to build intelligent systems*, 2<sup>nd</sup> ed., O'Reilly Media,
- Agarap, A. F. M. (2018). Statistical analysis on E-commerce review, with sentiment classification using bidirectional recurrent neural network(RNN). *arXiv*, 2018; arXiv:1805.03687
- Ali, F., El-Sappagn, S., & Kwak, D. (2019). Fuzzy ontology and LSTM-based text mining: a transportation network monitoring system for assisting travel. *Sensors*, 19, 234.
- Almiron-chamadoira, P.(2018). Online reviews as a genre: A semiotic analysis of Amazon.com 2010-2014 review on the categories 'Clothing' and 'Electronics', *DTUC'18*, October 3-5, 2018, Paris, France.
- Chong, A. Y. L., Li, B., Ngai, E. W. T., Ch'ng, E., & Lee, F.(2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: a big data architecture and neural network approach. *International Journal of operations & Production Management*, 36(4), 358-383.
- Duan, W., Gu, B., & Whinston, A. B.(2008). Do online reviews matter? -An empirical investigation of paenl data. *Decision Support Systems*, 45, 1007-1016.
- Elmurngi, E. I., & Gherbi, A.(2018). Unfair review detection on Amazon reviews using sentiment analysis with supervised learning techniques. *Journal of Computer Science*, 14(5), 714-726.
- Jia, S. (2018). Behind the ratings: Text mining of restaurant

- customers' online reviews, *International Journal of Market Research*, 60(6), 561-572.
- Kar, C., Lee, H., Tse, Y. K., Zhang, M., & Ma, J. (2019). Analysing online reviews to investigate customer behaviour in the sharing economy: the case of Airbnb. *Information Technology & People*, 33(3), 945-961.
- Kawaf, F., & Istanbuluoglu, D. (2019). Online fashion shopping paradox: the role of customer reviews and facebook marketing, *Journal of Retailing and Consumer Studies*, 48, 144-153.
- Kim, S., & Kim, Y-K. (2018). The interplay of regulatory focus and temporal distance on consumer evaluation of online reviews. *Clothing and Textiles Research Journal*, 36(3), 135-150.
- Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: a topic modeling application on airline passengers' online reviews. *Expert Systems With Applications*, 116, 472-486.
- Lee, C. K. H., Tse, Y. K., Zhang, M., & Ma, J. (2020). Analysing online reviews to investigate customer behavior in the sharing economy: The case of Airbnb, *Information Technology & People*, 33(3), 945-961.
- Lui, Y., Lu, J., Yang, J., & Mao, F. (2020). Sentiment analysis for e-commerce product review by deep learning model of Bert-BiGRU-Softmax, *Mathematical Biosciences and Engineering*, 17(6), 7819-7837.
- McKinney, E., & Shin, E. (2016). Exploring criteria consumers use in evaluating their online formal wear rental experience: A content analysis of online reviews. *Clothing and Textiles*



*Research Journal*, 34(4), 272-286.

Praveenraj, D. D. W., Subramani, K., Murugan, R. T., & Manikandan, S. (2021), Exploring consumer reviews for men's fashion accessories in online purchase platforms using sentiment analysis, *Revista Geintex*, 11(2), 676-685.

Reich, T., & Maglio, S. J. (2020). Featuring mistakes: the persuasive impact of purchase mistakes in online reviews. *Journal of Marketing*, 84(1), 52-65

Wang, W., & Chen, Q. (2020). Monitoring online reviews for reputation fraud campaigns. *Knowledge-Based Systems*, 195, 1-12.

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for e-commerce product review in Chinese based on sentiment lexicon and deep learning. *IEEE Access*, 8, 23522-23530.

Zhang, J., & Liu, R. R. (2018). The more the better? Exploring the effects of reviewer social networks on online review, *Journal of Marketing Management*, 35(17-18), 1667-1688.

# Abstract

## Classification and Prediction of Customer Reviews on Fashion Products in Online Shopping Mall (k-means clustering & LSTM)

Kim, Hyunsook

Seoul School of Integrated Sciences and Technologies

Advisor: Park, Cheong Yeul, Ph.D.

Online shopping for fashion products accounts for a high proportion of the transaction amount by product group. Product planning and inventory management is important for fashion product because there are many different types of products and individual preferences are highly differentiated. Online reviews have a direct impact on consumers' purchases, and are used as important indicators to identify their preferences and predict subsequent purchases.

Through analysis of fashion product reviews, various preferences and factors of satisfaction and dissatisfaction are revealed, and if these factors are used appropriately, it will be possible to increase consumer satisfaction and increase product planning and marketing efficiency. Nevertheless, different themes exist

simultaneously in one document and are composed of vast amounts of unstructured data with different expression styles and lengths. In order to effectively utilize them in practice, semantic analysis should be performed along with quantified the model.

The scope of this study is to create a sentiment prediction model by analyzing the review data of fashion products, and to categorize them for classification. For this purpose, 883,663 reviews from Amazon dataset for fashion products. were analyzed. For a score of 5 points, 1 and 2 were labeled as negative sentiment and 5 as positive sentiment, respectively, and a sentiment prediction model and classification model were built and verified using an artificial neural network algorithm.

The purpose of this study is to categorize online fashion reviews and create a sentiment prediction model in order to suggest online prediction and classification system. Specific research questions for this purpose are as follows. First, we create a model for predicting the sentiment of online fashion reviews and verify their performance. Second, we classify online fashion reviews and verify their performance.

In order to build a model for predicting sentiment for online fashion reviews, a model combined with various embeddings and algorithms was verified. Count Vectorization, TF-IDF, and Word2Vec were used for embedding, and the algorithm was verified using SVC, logistic regression, random forest classifier, bagging classifier, and multinomial NB. As a result of comparing these results with the bi-directional sequence LSTM model, the performance of the bi-directional LSTM model was the highest with an accuracy of 0.94. In addition, as a result of categorizing online fashion reviews with the k-means clustering algorithm, they were classified into three groups. Each group was named "Size & Fit", "Quality & Price", and "Appearance" followed by ANOVA mean analysis and post-hoc analysis. Significance was verified by silhouette score. In addition, as a

result of testing the bi-directional LSTM on three classified groups, the accuracy of all three groups was 0.93. To test the performance of the classification model, the overall evaluation text in dataset was labeled with three classification criteria and modeled as a bi-directional LSTM, resulting in an accuracy of 0.79.

In conclusion, this study identified factors that determine consumer satisfaction after purchase by proposing a model for classifying fashion product reviews and building a sentiment prediction system, and provided basic data necessary for review-based effective product planning and marketing strategies.. In addition, the rating prediction model of this study can be effectively used to predict the sentiment of review data when there is no rating, and it can be used for efficient review management, such as classifying fake ratings to prevent top exposure. Furthermore, if reviews are managed by category according to the review classification criteria of this study, it will be possible to respond more efficiently and quickly to consumer needs.

Key words: online review, fashion product, classification, k-means clustering, bidirectional LSTM, sentiment analysis

Student Number: 411005