

## Sales Forecasting Model Considering the Local Environment

Chul Soo Kim<sup>1,a</sup>, Su Min Oh<sup>a</sup>, So Yeon Park<sup>a</sup>

<sup>a</sup>Department of Computer Science and Statistics, Jeju National University

### Abstract

Today, local environmental factors has an influence on our society. Local enviromental factors, as well as weather-related natural phenomena, social phenomena are also included. In this paper, numeric factors and categorical factors were analyzed, looking for a local environmental factors affecting the company's sales. Sales model by performing a regression analysis based on this was implemented. Sales model considering the local environment had an accuracy of 88.89%.

Keywords: Sales forecasting model, local environmental factor, clustering analysis, k-means algorithms.

### 1. 서론

오늘날 기업에서 매출을 예측하는 것은 매우 중요한 관심사이며, 다양한 방법을 통해 제안되고 있다. 일반적으로 기업에서는 과거의 정보와 경험을 기반으로 예측하거나 경영자의 의도적인 전략에 의하여 제안되고 있다. 기업에서는 과거의 정보를 기반으로 미래의 매출을 전망하며, 대부분의 예측 모형에서는 이러한 정보를 모형의 요인으로 활용하고 있다. 또한, 경영자의 경험 및 전략에 의해 의도적으로 기업의 상황을 조절 할 수 있다. 즉, 제어가 가능한 요인을 기반으로 제안된다고 할 수 있으며, 이는 신뢰할 수 있는 예측을 가능하게 한다. 하지만 기존의 방법들은 제어가 불가능한 요인에 대한 논의는 다소 미흡한 편이다. 본 논문에서는 제어가 불가능한 요인들을 지역환경요인으로 판단하고 이를 적극적으로 활용하는 방법을 제안하고자 한다.

WMO (1964)에 의하면 기상에 대한 투자는 투자액 대비 10배 이상의 효과가 있다고 보고했다. 또한, 기상정보는 국가마다 차이가 있을 뿐만 아니라 국가 내에서도 지역에 따라 큰 차이를 보이고 있다. Jung (2006)는 모든 사업 분야의 70% 이상이 날씨로부터 직간접적인 영향을 받고 있으며 국가 GDP의 약 20~30%가 직간접적으로 연결돼 있어 날씨정보는 기업 경영에 매우 중요한 요소라고 하였다. 최근 날씨와 같은 환경요인을 적극적으로 활용하는 기업의 수가 증가하고 있다. 예를 들어, Lotte-World (2002), Pineresort (2010), CJ Home Shopping (2003) 등의 기업에서는 날씨 정보를 적극적으로 활용하여 매출을 향상시키고 있다.

Agn (2002)는 회귀모형을 이용해 원가예측을 시도하였으며, 수식 (1.1)에서는 내부적으로 발생하는 다양한 요인을 독립변수로 하고, 원가를 종속변수로 하는 회귀모형을 설정하였다.

$$Y_j = B_0 + B_1X_1 + B_2CPI + B_3MO, \quad (1.1)$$

$Y_j$  = 각 영업부문의 원가 예측치,  $B_0$  = 회귀상수,  $B_n$  = 회귀계수 ( $n = 1, 2, 3$ ),

$X_1$  = 각 영업부문의 매출액,  $CPI$  = 연도별 월별 소비자 물가지수,  $MO$  = 계절효과 터미변수.

<sup>1</sup> Corresponding author: Professor, Department of Computer Science and Statistics, Jeju University, 102 Jejudaehakno, Jeju Special Self-Governing 690-756, Korea. E-mail: kimcs@jejunu.ac.kr

Table 1: The measuring process of the local environmental factors's influence

Input: 기업의 매출 데이터, 지역의 환경요인
1 단계. 해당 지역의 환경요인
1-1. 각 요인별 상관분석을 통한 요인선정
2 단계. 선정된 요인검정
2-1. ANOVA 분석
3 단계. 선정된 요인에 대한 <i>k</i> -means 군집화
3-1. 각 군집을 이용한 회귀분석
3-2. 모형 분석
4 단계. 모형
4-1. 지역환경요인의 영향력
4-2. 지역환경요인을 고려한 매출 모형 구현
Output: 지역환경요인의 영향력을 이용한 매출 예측

이모형은 기업에서 발생하는 여러 요인을 이용하여 매출을 예측하는데 용이하다. 하지만, 기업의 내부적 요인으로 설계되어 있기 때문에 지역적으로 발생할 수 있는 다양한 요인들의 영향력을 설명하기에는 부족하다. 지역환경요인으로는 날씨가 대표적이라 할 수 있으며, 현재 날씨와 관련하여 많은 연구가 이루어지고 있고, 실제 날씨를 기업 경영에 적극 활용하는 사례가 많아지고 있다. Lee 등 (2005)는 국내에서 측정된 기후데이터를 저장 및 관리하기 위한 데이터를 모델링하고, *k*-means 알고리즘을 이용하여 국내 기후 데이터를 지역적 특성에 따라 군집함으로써 체계적인 데이터 정보를 제공하였다. Ahn (2007)은 날씨가 기업의 매출에 미치는 영향에 대하여 연구하였으며, Song (2007)은 제주지역의 관광발전을 위한 기상요인에 대한 연구를 진행하였다. Lee 등 (2005)는 마케팅 분야에 대한 날씨효과를 연구하였고, Jang (2010)은 호텔산업의 활성화를 위한 데이터 마이닝 기법 활용에 대한 연구를 진행하는 등 많은 연구가 이루어지고 있다. 본 논문에서는 기업 매출에 영향을 주는 지역 환경요인을 찾고, 이에 대한 다양한 분석을 통해 지역환경요인의 영향력을 합리적으로 평가하는 방법을 제안함으로써 미래의 지역경제 활성화에 도움이 되고자 한다. 매출과 관련된 다양한 요인들은 교차분석, ANOVA분석, 군집분석기법을 이용하였으며, 제안하는 매출 예측 모형에서는 교차분석과 ANOVA검정을 이용하여 범주형 데이터를 분석하였다. 수치형 데이터의 분석은 Bradley (1998)가 제안한 *k*-means 알고리즘의 군집분석 기법을 이용하여 분석하였고, 이를 회귀분석하여 모형을 구현하였다.

## 2. 본론

본 논문에서는 지역환경요인이 매출에 미치는 영향력을 측정하고 이를 이용하여 매출을 예측할 수 있는 모형을 제안한다. 먼저, 지역환경요인이라는 것은 어느 지역에서 발생할 수 있는 공통적 요인과 더불어 타 지역과 구별되는 해당 지역만의 고유한 요인을 포함한다. 예를 들어, ‘주말이 매출에 미치는 영향’, ‘시간대별 매출의 변화’는 주말과 시간이 매출의 요인이 되며, 이러한 요인들은 지역에 상관없이 어느 지역이나 해당되는 공통된 요인이라고 할 수 있다. 하지만 날씨나 오일장과 같은 환경요인들은 그 지역의 고유한 특성으로서 해당 지역에만 영향을 보인다고 할 수 있다. 제주도의 오일장과 같은 요인들은 지역의 고유한 특성이며, 이 요인들은 그 지역의 기업 및 외식업체 등에 영향을 미칠 것이다. 본 논문에서는 지역환경요인을 선정하는 방법과 이를 이용하여 요인의 영향력을 측정하고, 매출을 예측하는 모형을 구현하였다.

Table 1은 지역환경요인의 영향력을 측정하는 과정을 보여준다. 1단계에서는 기업의 매출 데이터를 수집하고 매출에 영향을 줄 것으로 예측되는 다양한 지역환경요인 변수를 조사한다. 예를 들어 날

씨 데이터의 경우 해당 업체가 위치한 지역의 기온, 강수량과 같은 수치형 변수와 날씨정보(맑음, 흐림, 비)와 같은 범주형 변수를 포함한다. 또한, 오일장 데이터의 경우 해당 업체가 위치한 지역의 날씨를 이용하였으며, 이후 분석을 용이하게 하기 위하여 오일장인 날과 오일장이 아닌 날로 간소화 하였다. 지역환경요인은 매출과의 상관분석을 통하여 선정한다. 선정된 지역환경요인들은 요인의 속성에 따라 분석 방법을 구분한다. 2단계에서는 선정된 요인 중 범주형 요인에 대하여 ANOVA분석을 한다. 오일장유무에 따른 매출의 유의성, 날씨환경(맑음, 흐림, 비)에 따른 매출의 유의성, 주말유무에 따른 매출의 유의성 등을 검정한다. 또한, 이러한 요인들이 독립적으로 발생하므로 범주속성의 영향력은 각각의 곱으로 표현 할 수 있다. 본 논문에서는 지역환경요인들 중 기온요인을 주요한 요인으로 고려하고 있으며, 기온 데이터는 수치형 데이터로 이루어져 있다. 3단계에서는 수치형 데이터의 분석에 잘 알려진  $k$ -means 알고리즘을 이용하여 군집분석을 수행한다.  $k$ -means 알고리즘은 데이터의 outlier와 군집 수에 따라 분석의 결과가 다르게 나타나며,  $k$ -means 알고리즘을 적용하기 위해 outlier 제거 및 최적의 군집수를 결정해야 한다. Outlier는 분석 결과를 왜곡시키며, 데이터의 분산을 증가시키는 주요한 요인이 된다. Hautamaki 등 (2005)는 Outlier 제거 방법을 소개하였으며, i) distribution based, ii) depth based, iii) distance based, iv) clustering based, v) density based로 구분하였다. 먼저, Distribution based는 데이터가 통계적인 분포를 따른다는 것을 이용하여 outlier를 제거하는 방법이다. 예를 들어, 데이터가 정규분포를 따른다면, 데이터의  $\pm 3\sigma$  범위 밖의 데이터는 outlier로 간주한다. Depth based는 convex hull를 이용하여 hull에 해당하는 데이터를 outlier로 간주한다. Distance based는 데이터간의 유클리드 거리를 기반으로 자신을 제외한 데이터들로부터 거리가 가장 멀리 있는 데이터를 outlier로 한다. Clustering based는 데이터의 패턴이 상이한 데이터를 outlier로 판단하는 기법이며, 수치형 데이터뿐만 아니라 범주형 데이터의 분석에 유용하다. Density based는 데이터를 연결하는 그래프를 이용하는 방법으로 Mutual k-Nearest Neighbor(MkNN)에 의한 Brito 등이 알려져 있다. 본 연구에서는 기업의 매출은 특정요인을 독립변수로 하는 회귀모형으로 표현할 수 있으며, 이는 특정요인에서 관측된 매출의 분포는 정규분포를 따른다고 가정한다. 본 논문에서는 distribution based 방법을 이용하여 outlier를 제거하고  $k$ -means 알고리즘을 적용하였다.

$$\widehat{Y}_K = \bar{Y}_i \cdot C_r + \bar{Y} \quad (2.1)$$

$$Y_K = \beta_2 E^2 + \beta_1 E + \beta_0 = \beta_2 \left( E + \frac{\beta_1}{2\beta_2} \right)^2 + \frac{4\beta_0\beta_2 - \beta_1^2}{4\beta_2}, \quad \bar{Y} = \beta_2 E^2, \quad (2.2)$$

$$\widehat{Y}_K : \text{군집수가 } K \text{ 일 때의 예상 매출}, \quad \bar{Y}_i : \text{군집분석결과 } K_i \text{ 번째 군집의 평균}, \quad (2.3)$$

$$C_r : r \text{ 번째 범주형 요인들의 비율}, \quad Y_K : \text{군집분석결과에 대한 회귀식}, \quad (2.4)$$

$$E : \text{지역환경요인의 대표요인}, \quad \bar{Y} : \text{지역환경요인으로 인해 발생하는 영향}. \quad (2.5)$$

수식 (2.1)은 과거의 매출정보로부터 예측된  $\bar{Y}_i$ 에 대하여 제어할 수 없는 지역환경요인( $\bar{Y}$ )을 고려한 모형이다.  $\bar{Y}_i$ 는 과거의 매출정보로부터 예측된 군집( $K_i$ )의 평균이며, 지역환경요인의 대표요인을 독립변수로 하게 된다. 결국, 대표요인의 특성에 따라 회귀모형은 다양하게 나타날 수 있다. 본 논문에서는 기온을 대표요인으로 하였으며, 기온이 겨울에는 내려가고 여름에는 상승한다는 것을 고려하여 2차 식으로 구현하였다. 즉, 수식 (2.1)에서 지역환경요인( $E$ )이 주어질 경우,  $\beta_2$  만큼 기존에 알려진 매출  $\bar{Y}_i$ 에 영향을 주게 되고, 이 영향력은 회귀식을 원점에 대하여 표현하면 지역환경요인의 영향력( $\bar{Y}$ )이 된다. 결국, 지역환경요인을 고려한  $\widehat{Y}_K$  예상매출의 범위는 다음과 같다.

$$\left[ (\bar{Y}_i \cdot C_r - \bar{Y}) \cdot \theta, (\bar{Y}_i \cdot C_r + \bar{Y}) \cdot \theta \right] \quad (2.6)$$

수식 (2.6)은 매출의 추정 범위이다. 과거 매출 데이터를 기반으로 하는 군집의 평균매출  $\bar{Y}_i$ 에 대하여

Table 2: Classification according to the characteristics of factors

매출요인	일반요인	지역환경요인
건당주문금액	월 구분	기온, 강수량
주문건수	일 구분	날씨정보
일 매출	요일 구분	오일장정보
LN(일 매출)	주말 구분	절대기온편차

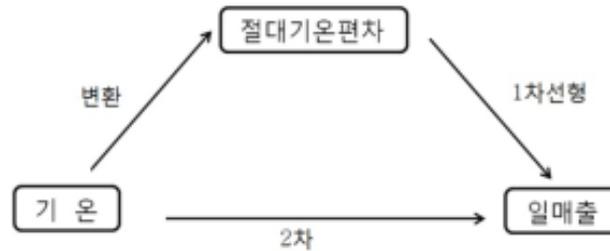


Figure 1: The relation between temperature and daily sales

범주형요인으로 구성된 비율  $C_r$ 를 적용하며, 이에 대하여 지역환경요인의 영향력과 경영자의 매출변화 예측치를 가중치(weight  $\theta$ )를 기반으로 하는 매출범위를 예측한다.

### 3. 실험

#### 3.1. 자료

본 논문에서는 환경요인의 영향력 평가를 위하여 2010년 7월 1일부터 2011년 6월 30일까지 1년 동안의 제주특별자치도 서귀포시 지역에 위치한 소규모 피자외식업체를 선정하여 매출데이터를 조사하였으며, 2011년 7월 1일부터 2011년 8월 31일까지의 2달간 데이터를 모형의 검증에 이용하였다. 기온, 강수량, 날씨정보는 기상청자료를 활용하였으며, 오일장과, 주말에 대한 구분은 날씨만으로 구분할 수 있다.

Table 2는 조사된 매출데이터의 요인별 구분을 보여주고 있다. 매출요인으로는 건당주문금액, 주문건수, 일 매출이 있다. 지역환경요인 중 공통적인 요인으로는 월 구분, 일 구분, 요일구분, 주말 구분으로 하였으며 고유한 요인으로는 조사된 지역의 기온, 강수량, 날씨정보, 오일장정보를 사용하였다. 기온에 대해서는기온편차의 절댓값( $|\text{평균기온} - \text{기온}|$ )을 ‘절대기온편차’로 변수를 추가하였으며, 매출에 관련된 요인에 대해서는 일 매출을 로그 변환한 변수를 추가하였다. 일 매출에 대한 로그변환 변수는 데이터의 outlier를 제거하고,  $k$ -means 알고리즘의 군집분석에 활용된다. Figure 1은 기온과 일 매출의 관계를 보여주고 있다. 기온은 일 매출과 단순회귀분석에서는 유의하지 않으며, 기온의 특성상 2차 회귀식으로 잘 적합 한다.

#### 3.2. 범주형 데이터

##### 3.2.1. 범주형 요인에 대한 분석

선정된 지역환경요인들 중 범주형 데이터는 오일장변수(오일장, 평일), 주말변수(월~금, 토일), 날씨변수(맑음, 흐림, 비)이다. 이 요인들에 대하여 ANOVA분석을 수행한다.

Table 3은 요인별 일매출의 평균과 ANOVA검정 결과이다. 오일장변수와 주말변수는 매출에 대하여 유의하게 나타났다. 하지만, 날씨변수는 매출에 대하여 유의하다고 할 수 없다. 본 논문에서는 각

Table 3: The Results of the ANOVA between the local environmental factors and sales

요인	구분	매출평균	표준편차	df	F	p-value
오일장변수	평일	138377	55712	1	4.54	0.0338
	오일장	123082	59208			
주말변수	월~금	122674	50822	1	47.49	0.00001
	토, 일	165630	58546			
날씨변수	맑음	1333600	54156	2	0.12	0.8876
	흐림	136857	56514			
	비	135855	60011			

Table 4: The influence( $C_r$ ) of the local environmental factors

구분 (r)	오일장변수		주말변수		날씨변수		일매출	$(C_r)$
	구분	상대비율	구분	상대비율	구분	상대비율		
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00
000	평일(0)	1.00	월-금(0)	1.00	맑음(0)	1.00	142938.9	1.00

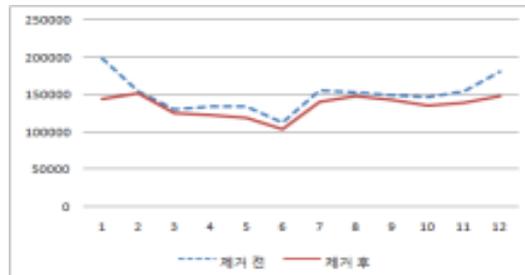


Figure 2: Sales comparisons before and after removing outliers

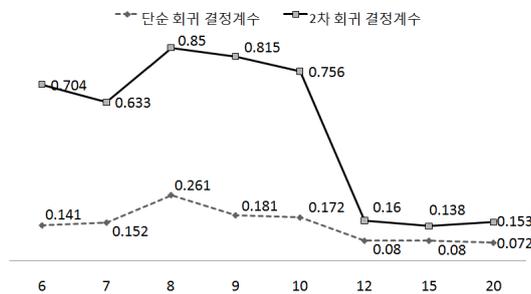
변수에 대하여 수준에 대한 매출비를 활용하게 되며, 오일장변수와 주말변수는 각 수준에 대하여 매출 비율이 날씨변수의 수준에 따른 변화보다 크게 나타나게 된다.

### 3.2.2. 범주형 요인의 영향력( $C_r$ )

Table 4는 범주형 요인에 대한 영향력( $C_r$ )을 보여주고 있다. 오일장변수의 상대비율은 평일과 오일장의 매출의 비로 계산하였으며, 주말변수의 상대비율은 월~금(평일)과 토, 일(주말)로 구분했을 경우의 매출 비가 된다. 날씨변수의 경우 날씨의 구분을 맑음, 흐림, 비로 구분하여 각각의 날씨에 대한 매출의 비로 하였다. 이는 Table 3에서의 ANOVA 분석결과와 연관이 있으며, 유의한 요인일수록 상대비율의 차이가 크게 나타나고 있음을 알 수 있다.

Table 5: ( $R^2$ ) and  $p$ -value by the number of clusters

K	기온, LN(일매출), 주말변수, 오일장변수			
	단순회귀모형		2차다항회귀모형	
	$R^2_K$	$p$ -value	$R^2_K$	$p$ -value
6	0.141	0.463	0.704	0.161
7	0.152	0.388	0.633	0.134
8	0.261	0.196	0.850	0.009
9	0.181	0.254	0.815	0.006
10	0.172	0.234	0.756	0.007
12	0.080	0.373	0.160	0.455
15	0.080	0.307	0.138	0.409
20	0.072	0.054	0.153	0.243

Figure 3:  $R^2$  by the number of cluster

### 3.3. 수치형 데이터

#### 3.3.1. 수치형 데이터의 영향력

수치형 데이터의 지역환경요인들 중 수치형 데이터의 영향력을 측정하기 위하여  $k$ -means 알고리즘을 수행하였다. 알고리즘을 적용하기 위하여 매출데이터의 outlier를 제거하였다. 본 논문에서는 건당 주문금액에 대해 평균 주문금액의  $\pm 2\sigma$  (95.5%) 범위 밖의 매출에 대해서는 정상적이지 않은 매출로 판단하였다. Figure 2는 매출에 대한 outlier를 제거한 전후 매출 평균의 비교 그래프이다.

#### 3.3.2. $k$ -means 알고리즘을 이용한 군집분석

$k$ -means 알고리즘에서 군집의 개수( $K$ )를 정하는 것은 매우 중요하다. 본 연구에서는 최적의 군집 개수를 결정하기 위하여 군집 수에 따른 회귀분석의 결정계수 값을 활용하였다. 즉,  $K$ 개로 군집으로 이루어진 군집의 대푯값에 대한 회귀분석으로 구한 결정계수를 이용한다. Table 5는 군집의 개수별 지역환경요인의 변화에 따른 군집과 매출간의 단순 회귀모형의 결정계수와 2차 다항회귀모형의 결정계수이다. 군집의 개수를 8개로 하였을 때, 2차 다항회귀모형에서 결정계수가 0.850으로 가장 높으며 유의하게 나타났다. Figure 3의 그래프는 매출에 대한 지역환경요인의 군집 개수별 단순 선형회귀모형과 2차 다항회귀모형의 결정계수를 비교하고 있다. Table 6은 군집의 수를 8개로 하였을 때의 군집의 평균 매출과 빈도를 보여주고 있으며, Figure 4는 실제 데이터의 기온과 매출간의 산점도와 군집분석결과에 따른 산점도를 보여주고 있다 Table 7은 군집분석결과( $k = 8$ )에 대한 회귀분석 결과이다. 실험 대상 업체의 지역환경요인과 매출과의 회귀모형은 유의하게 나타나고 있으며, 군집분석결과에 대한 회귀식

Table 6: The average temperature and frequency on cluster analysis

$K_i$	평균기온	매출	빈도
1	-0.9	207933.3	12
2	4.4	169591.7	36
3	10.1	135053.1	49
4	14.3	149704.8	63
5	18.5	150175.7	37
6	22.4	131453.7	54
7	26.1	146866.7	33
8	29.7	156208.2	49

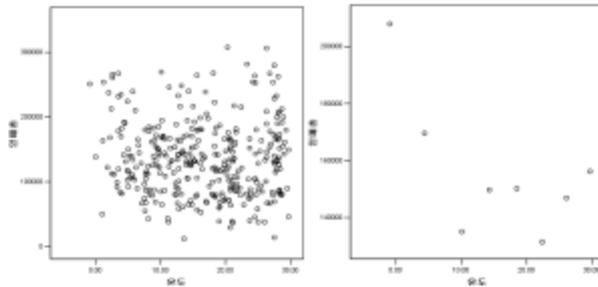


Figure 4: Plot of temperature vs daily sales and the cluster analysis result plot

Table 7: Regression analysis results table on the cluster analysis

Dependent	Mth	Rsqr	d.f	F	p	$\beta_0$	$\beta_1$	$\beta_2$
평균매출	LIN	.418	6	4.30	.083	178697	-1465.4	
평균매출	QUA	.860	5	15.35	.007	197429	-6407.7	170.433

은 다음과 같다.

$$Y_K = \beta_2 E^2 + \beta_1 E^2 + \beta_0 \approx 170(E - 18.84)^2 + 137062, \quad (3.1)$$

$$\widehat{Y}_K = \bar{Y}_i \cdot C_r + 170E^2. \quad (3.2)$$

수식 (3.1)는 내부요인에 의한 매출에 대한 군집분석 결과를 이용하여 구현한 예측 모형이며, 모형을 이용한 매출의 최대기대치와 최저기대치는  $\bar{Y}_i \cdot C_r \pm \bar{Y}$ 이다.

Figure 5는 군집에 대한 회귀 그래프와 잔차도표이다. 왼쪽 그래프는 단순 선형회귀모형과 2차 다항회귀모형의 그래프를 보여주고 있으며, 오른쪽 산점도 도표는 2차 다항회귀모형에 대한 잔차도표이다. 결론적으로, 예상매출( $\widehat{Y}_K$ )의 범위는 다음과 같다.

$$\left[ (\bar{Y}_i \cdot C_r - 170E^2) \cdot \theta, (\bar{Y}_i \cdot C_r + 170E^2) \cdot \theta \right]. \quad (3.3)$$

### 3.3.3. 모형 검증을 위한 weight $\theta$

본 논문에서는 제안하는 예측모형의 정확도를 검증하기 위해 해당업체의 최근 데이터를 조사하여 검증하였다. 조사 데이터 중 나타나고 있지 않는 날짜는 영업을 하지 않거나, 정상적인 영업을 이루어지지 않은 날이다. 이에 대한 제거 방법은 일 매출이 정규분포를 따르는지 여부와 실제 업체로부터 조사된 결과를 이용하였다. 또한, 경영자는 전년도 대비 직원의 감축 및 경기 하락으로 인해  $\theta = 40\%$  정

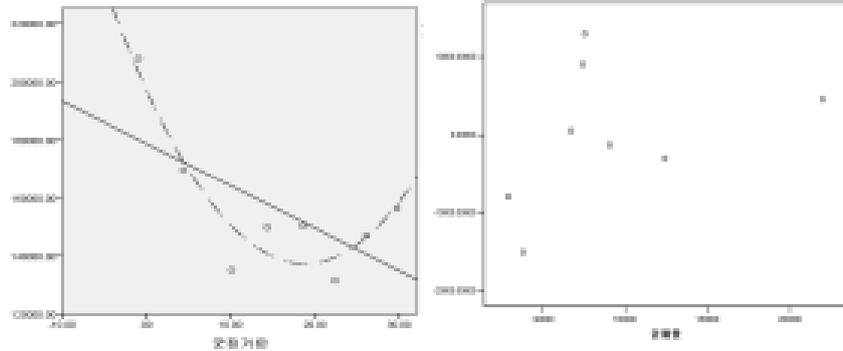


Figure 5: Estimated regression models and residual plot of the linear regression model

Table 8: Sales predictions results by the proposed model

날씨	구분	$(\bar{Y}_i \cdot C_r)$	실제매출( $y_i$ )	$\hat{Y}_K$ 하한	$\hat{Y}_K$ 상한	1 = 성공, 0 = 실패
7/2	011	202,234.5	134,700	116,988	206,587	1
7/6	002	148,334.7	100,200	85,769	151,566	1
7/8	002	148,334.7	163,700	82,339	154,997	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
8/20	012	200,251.8	96,100	113,489	206,914	0
8/23	002	148,334.7	108,300	84,297	153,038	1
8/25	000	146,866	117,700	80,351	154,635	1

Table 9: Accuracy(%) of prediction model according to the weight( $\theta$ )

weight	정확도	weight	정확도	weight	정확도
$\theta$	%	$\theta$	%	$\theta$	%
1.00	7.41	0.65	66.67	0.30	88.89
0.95	18.52	0.60	77.78	0.25	88.89
0.90	29.63	0.55	85.19	0.20	88.89
0.85	29.63	0.50	88.89	0.15	88.89
0.80	40.74	0.45	88.89	0.10	88.89
0.75	48.15	0.40	88.89	0.05	88.89
0.70	59.25	0.35	88.89	0.00	88.89

도의 매출 하락을 예상하고 있으며, 이를 경영자의 예측(weight)으로 적용하였다. 즉, weight( $\theta$ )는 기업의 경영자가 예상하는 매출의 변화이다. 또한, 업체의 특성상 피자판매는 하루에 만들 수 있는 수가 고정적이므로 예측매출의 상한에는 weight를 적용하지 않고, 예측매출의 하한에 대해서는 weight를 적용하였다.

Table 8은 제안하는 예측모형을 통해 예측한 결과이다. 모형을 통해 매출을 예측이 성공한 경우는 1로 표시하였으며, 실패한 경우는 0으로 표현하였다. Table 9는  $\theta$ 에 따른 예측모형의 정확도(%)이며, Figure 6은 정확도의 변화추이이다.  $\theta = 0.50$ 일 때 가장 높은 정확도를 보이고 있으며, 경영자가 40%의 매출하락을 예상하였을 때 77.78%의 정확도를 보였다. Table 9는  $\theta$ 가 0.5일 때 제안하는 모형의 정확도가 가장 높게 나타났으며, 이는 실제 매출이 40%가 아닌 50%의 매출 하락을 의미한다. 본 연구에서는 실제 매출의 88.89%를 예측하였다.

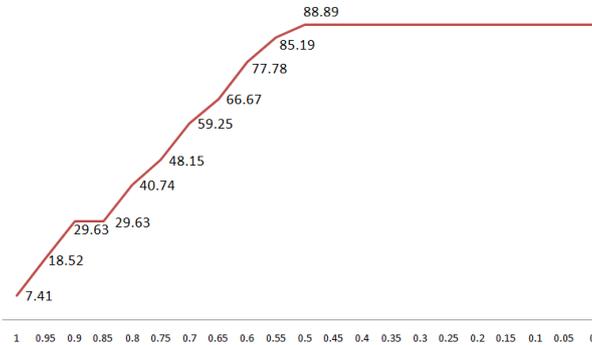


Figure 6: The accuracy of prediction model

#### 4. 결론 및 제언

기업에서는 다양한 요인과 방법을 통해 매출을 예측하고 있다. 미래에 대한 예측은 과거의 경험적 정보를 통해 결정되고 있다. 이는 매우 신뢰할 수 있는 방법이며 지금까지 우수한 결과를 보여주고 있으나, 기후의 변화, 지역 내에서 발생하는 외부현상들에 대한 활용은 아직 미흡한편이다. 최근 날씨와 같은 외부적인 환경요인들이 기업 경영에 큰 영향을 주고 있다. 본 논문에서는 날씨와 같은 자연현상과 지역내에서 발생하는 다양한 현상을 지역환경요인으로 판단하고, 이를 이용하여 매출을 예측할 수 있는 방법을 제안하였다. 지역환경요인들은 범주형 데이터와 수치형 데이터로 구분하여 ANOVA 분석과  $k$ -means 알고리즘을 통해 분석하고, 모형을 제안하였다. 지역환경요인들 중 오일장과 같은 경제현상, 주말에 대한 구분, 날씨구분은 범주형 데이터로 판단하고, 각 요인의 비율을 적용하였으며, 기온과 기업의 일 매출은 수치형 변수로 판단하여 군집분석 하였으며, 이를 기반으로 회귀분석을 수행하였다. 이 논문에서 제안하는 모형에서는 경영자의 예측과 지역환경요인을 이용하여 88.89%의 매출을 예측하였다. 하지만 매출에 영향을 주는 가중치(weight)를 경영자의 주관적 예측으로 적용하였고, 이는 차후 연구를 통해 경제활동지표 등을 활용하는, 좀 더 객관적인 방법이 필요 할 것이다. 지역환경요인의 변화로 인한 매출의 영향력에 대한 연구는 아직 미흡하지만, 향후 미래 사회에서는 예측모형의 주요한 요인으로 자리잡을 것으로 기대된다. 그러므로 우리는 지역적으로 발생하는 환경요인에 보다 다양한 관심과 연구가 필요하며 이를 적극적으로 활용할 수 있는 많은 연구가 있어야 할 것이다.

#### References

- Ahn, K.-H. (2007). *The Effect of Weather on Firm's Sales*, Sejong university.
- An, K.-H. (2002). An empirical study on formulating hotel sales forecasting models, *Korea Journal of Tourism Research*, **17**, 43–57.
- Bradley (1998). Refining initial points for clustering, *Proc, 15th Internat. Conf. on Machine Learning*, Morgan Kaufmann, Los Altos, CA.
- CJ Home Shopping (2003). Surplus achieved eight years by Weather, seasonal specialized products. Korea Meteorological Agency, 2002, weather management Grand prize
- Hautamaki, V., Cherednichenko, S., Karkkainen, I., Kinnunen, T. and Franti, P. (2005). *Improving K-Means by Outlier Removal*, Springer-Verlag Berlin Heidelberg.
- Hotel-Lotte Lotte-World (2002). Weather Management- Success factors of the management of the world's largest indoor theme park, Lotte World. Korea Meteorological Agency, 2002, weather management Grand prize

- Jang, Y.-W. (2010). A study of weather marketing a hotel industry through data mining technique, Sejong university.
- Jung, Y.-M. (2006). Weather information is money; This practices. *Samsung Global Environment Research Center*.
- Lee, Y.-K., Kim, W.-T., Jung, Y.-J., Kim, K.-D. and Ryu, K.-H. (2005). Cluster analysis of climate data for applying weather marketing, *Korea Spatial Information Society*, 7, 33–44.
- Pineresort (2010). Expanding the utilization rate of Condo and Sky resort by the prior notification of weather information. Korea Meteorological Agency, 2004, weather revenue award
- Song, S.-S. (2007). *The Impact of Weather Factors on Tourism-With Focus on Jeju Island*, Kyonggi university.
- WMO (1964). Weather and man, WMO-No. 143 TP. 67 (*Geneva: World Meteorological Organization*).

2012년 8월 27일 접수; 2012년 10월 6일 수정; 2012년 11월 5일 채택