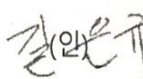




## 연구논문/작품 제안서

2018 년도 제 1 학기

논문/작품	○논문( <input checked="" type="checkbox"/> ) ○작품( <input type="checkbox"/> ) ※ 해당란에 체크
제목	지역의 성장정도에 따른 비만을 변화 예측에 관한 연구
GitHub URL	<a href="https://github.com/kasta12/graduationPaper.git">https://github.com/kasta12/graduationPaper.git</a>
팀원명단	길 은 규  (학번: 2012314166 )

2018 년 3 월 23 일

지도교수 : 김 응 모

서명

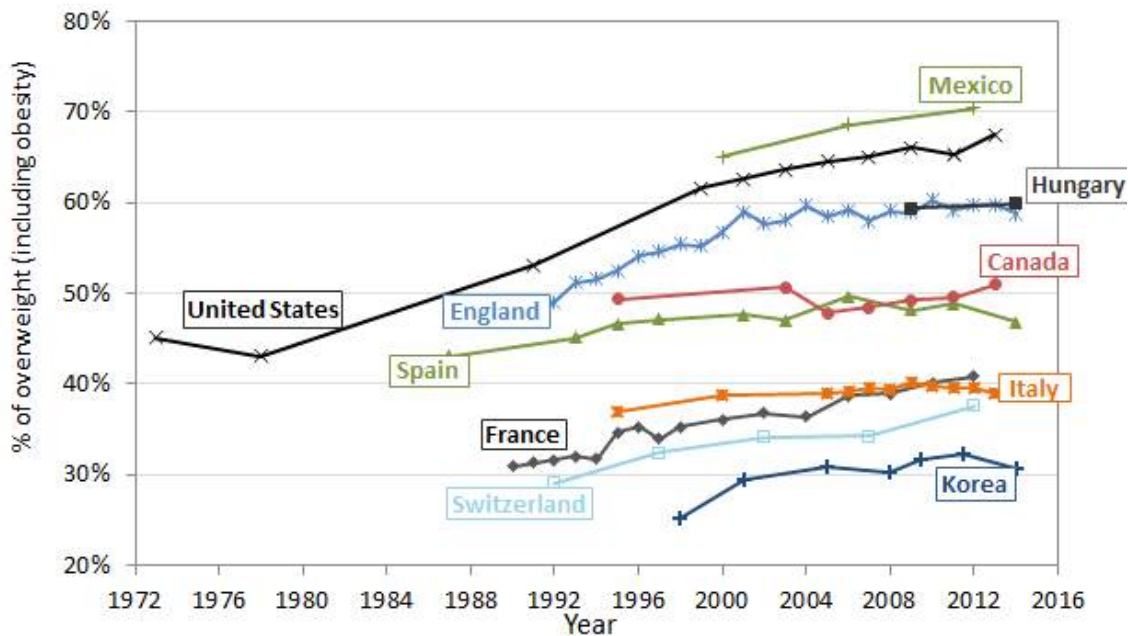


## 1. 과제의 필요성

### 1.1 연구 배경

21세기가 되면서 세계는 이전과 비교 할 수 없을 정도로 많은 분야에서 발전을 거듭해 왔다. 특히 과학기술이 이전과는 비교할 수 없을 정도로 발달함에 따라 여러 다른 분야에서의 발전을 촉진 시켰다. 이에 따라 이전에는 없던 많은 일자리가 생겨났고 사람들의 삶은 점점 더 풍족해 졌다.

이 변화는 우리나라에도 똑같이 적용되었다. 과거의 우리나라가 먹고 살기에 급급했다면 21세기에 들어서면서 먹을 음식이 없어 굶는 경우는 거의 보기 힘들어 졌다. 사람들은 먹고사는 문제가 해결되니 사치품이나 여가 및 문화생활에 관심을 가지게 되며 이전과는 비교할 수 없는 행복한 삶을 누릴 수 있게 되었다[1].



[그림 1] 주요 OECD 국가들의 연도별 과체중률(비만 포함) 변화

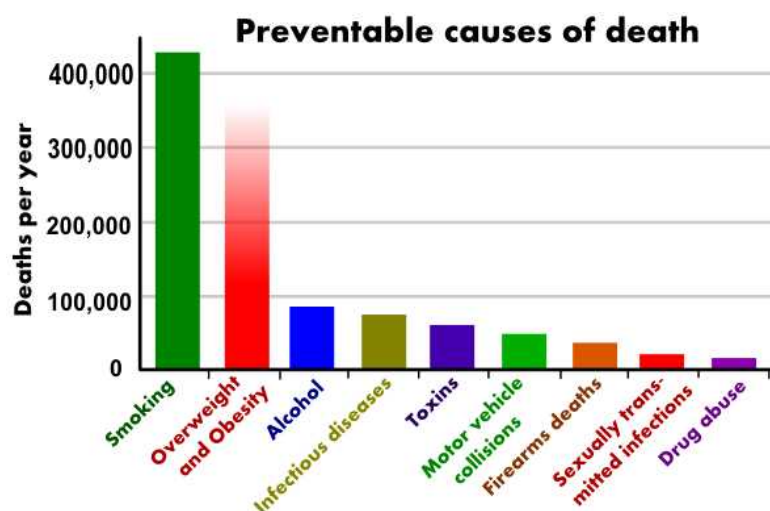
하지만 이러한 발전이 긍정적인 변화만을 초래한 것은 아니다. 사람들이 경제적으로 편하고 안정적인 삶을 누리게 됨에 따라 기존에 필요한 것보다 많은 것을 요구하게 되었다. 특히 이 변화는 '먹는 것' 에서 두드러져왔다. 이전과는 달리 삶이 풍족해 짐에 따라 필요이상의 음식을 섭취하게 되었다. 이는 광고 또는 부모, 친구등을 통한 사회적인 원인에 의해서 나타나기도 한다. 즉, 인간의 기본적인 욕구중 하나인 살기위해 먹는 것이 아니라 그것을 뛰어넘어 정신적인 만족 또는 다른 사회적 원인에 의해 필요 이상으로 음식을 먹는 시대가 온 것이었다 [2]. 이는 사람들의 체형과 건강에 직접적인 영향을 주게 되었다. [그림 1]에서 사

람들의 생활이 풍족해 지면서 과체중률이 늘어나는 것을 확인 할 수 있다[3]. 이 변화는 서구문화권에서 두드러진다. 근대화가 우리나라보다 더 빨리 일어난 서구 문화권 국가들의 사람들은 더 빨리 바쁘지만 풍족한 삶을 영유하게 되었고 이에 따라 급격한 체형 변화를 맞이하게 되었다. 실제로 지금도 서구문화권 국가의 길 거리를 다녀보면 우리나라와는 많이 다른 모습을 볼 수 있다.

물론 이 변화는 우리나라도 피해 갈 수 없었다. 비록 상대적으로 늦었지만 상당한 수준의 경제성장을 이루게 되었고 이에 따라 먹는 음식의 양도 달라졌으며 직접적인 체형변화를 초래했다. 이러한 점에 비추어 경제성장과 사람들의 건강 및 체형변화에 관심을 갖게 되었고 이 둘 사이의 관계에 대해 심도 있게 연구해보고자 한다.

## 1.2 연구 목적

국제 사회에서 이미 비만은 심각하게 다루어지는 문제점들 중 하나이다. 물론 옛날에는 없어서 못 먹었다지만 이제는 오히려 그 반대이다. 사람들이 필요 이상의 음식을 섭취함으로써 평균 체중은 점점 늘어나고 고도비만에 해당되는 사람들도 급격하게 증가했다. 이 극단에 해당하는 사람들은 일반 사람들과는 전혀 다른 외형적 변화를 겪게 된다. 하지만 더 큰 문제는 외형적 변화에 그치는 것이 아니라 이 증상이 다양한 합병증을 유발하여 건강을 위협한다는 점이다. 이 합병증들은 건강에 굉장히 치명적이며 심각한 경우에는 목숨을 위협하기도 한다.



[그림 2] 미국의 연간 예방 가능한 사망 원인별 사망자 수

[그림 2]에서 볼 수 있다시피 비만은 여러 사망원인들을 제치고 사망원인들 중 상위권에 위치해 있다[4]. 이처럼 비만은 절대 가볍게 다루어야 할 문제가 아닌 사람의 목숨과 직결되는 중요한 문제임이 틀림없다. 따라서 이 논문에서는 특정

지역 사람들의 비만율과 지역 성장정도를 비교하여 이 둘 사이에 정확히 어떤 상관관계가 존재하는지를 알아보고자 한다. 둘 사이의 관계를 명확히 하여 정확히 어떤 요소가 영향을 미치는 지 알 수 있다면 이 문제를 해결하는데 도움이 될 것이라 생각한다.

## 2. 선행연구 및 기술현황

### 2.1 선행연구

[그림 1]에서 확인 할 수 있드시피 우리나라를 제외한 20세기부터 일찌감치 높은 수준의 경제성장을 이룬 여러 국가 국민들의 과체중률을 살펴보면 우리나라보다 압도적으로 높은 것을 볼 수 있다[1]. 물론 이것은 경제성장에만 비례하는 것이 아니라 그 국가의 행정 서비스, 사회기반시설, 생활환경 등 다양한 사회적 측면이 종합적으로 반영된 것이다[5]. 따라서 이번 논문에서는 경제적 성장뿐만 아니라 다각적으로 이러한 요소들과 비만율과의 관계를 분석 할 것이다.

### 2.2 기술현황

주제가 데이터 분석인 만큼 논문 작성에 사용 될 기술들은 주로 데이터 분석에 이용되는 기술들이 될 것이다. 이에 이용될 기술들은 다음과 같다.

#### 2.2.1 머신러닝(Machine Learning)

머신 러닝은 이름에서 알 수 있듯이 기계를 즉, 컴퓨터를 인간처럼 학습시켜 스스로 규칙을 형성할 수 있지 않을까 하는 시도에서 비롯된 방법이다. 주로 통계적인 접근방법을 사용하며 인간이 하는 추론 방식과 매우 유사하며 강력하다. 카네기 멜런 대학교 교수 Tom M. Mitchell은 자신의 책'Machine Learning'에서 기계학습을 어떠한 작업(Task)에 대해 꾸준한 경험(Experience)을 통하여 그 T에 대한 성능(Performance)을 높이는 것이라고 정의했다[6]. 즉, 데이터를 의미하는 E가 가장 중요하며 좋은 품질의 데이터를 많이 가지고 있다면 보다 높은 성능을 끌어낼 수 있다는 것을 의미한다.

#### 2.2.2 회귀분석(Regression Analysis)

통계학에서 관찰된 연속형 변수들에 대해 두 변수 사이의 모형을 구한 뒤 적합도를 측정해 내는 분석방법이다. 특히 비선형 회귀 분석 방법은 공학에서 뿐만 아니라 수학 또는 사회과학 분야에서 널리 사용되는 기법이다[7]. 회귀 분석은 시간에 따라 변화하는 데이터나 어떤 영향, 가설적 실험, 인과 관계의 모델링 등의 통계적 예측에 이용될 수 있다.

### 2.2.3 K-평균 군집분석(K-means Clustering)

과거 엄청난 과학의 발전에 따라 과학 데이터 분석을 위한 여러 현대 기술이 개발됨으로서 과학을 포함한 다양한 분야에서 엄청난 양의 데이터가 축적되게 되었다. 이전의 데이터 추출 방법들은 이 방대한 데이터를 다루는데 부적합했다. 이에 따라 대두된 대표적 군집분석 방법들 중 하나인 K-평균 군집분석 방식은 현재 여러 분야에서 널리 적용되어지고 있다[8]. 이는 자료에 있는 각 개체를 유사한 특성을 지니는 K개의 그룹으로 분할하는 방법으로, 각 군집에 속하는 개체들의 평균값을 중심으로 하여 근접한 거리에 있는 개체를 묶어서 분할하게 된다. K-평균 군집분석은 거의 모든 형태의 자료에 적용이 가능하고 특별한 변환이 필요 없어서 적용이 쉽다는 장점이 있어 많은 분야에서 사용되고 있다. 이를 위해서는 사전에 정해져야 할 군집의 개수, K에 대한 정보가 주어져야 하고, K개 군집의 중심인 초기값(seed)들을 신중하게 결정하여야 한다.

### 2.2.4 워드 클라우드(Word Cloud)

태그 클라우드 또는 워드 클라우드는 메타 데이터에서 얻어진 태그들을 분석하여 중요도나 인기도 등을 고려하여 시각적으로 늘어놓아 웹 사이트에 표시하는 것을 말한다. 보통은 2차원의 표와 같은 형태로 태그들이 배치되며 이때 순서는 알파벳/가나다순으로 배치된다. 시각적인 중요도를 강조하기 위해 각 태그들은 그 중요도(혹은 인기도)에 따라 글자의 색상이나 굵기 등 형태가 변한다. 사용자는 이렇게 표시된 태그 중 마음에 드는 키워드를 발견하면 그것을 선택하여 그 메타 데이터에 원래 연결된 웹 페이지로 이동하게 된다.

### 3. 작품/논문 전체 진행계획 및 구성

[표 1] 논문 전체 진행계획표

월 내용	3	4	5	6	7	8	9	10	11
논문 제안서&서약서	——								
공공 데이터 수집	——	——							
데이터 분석 기법 연구		——	——	——					
머신러닝기반 분석법 연구		——	——	——					
데이터 구체적 분석			——	——	——				
논문 작성 및 학회 제출					——	——			
학회 발표자료 작성 및 제출						——	——		
논문 중간보고서						——	——	——	
학회								——	
논문 최종보고서								——	——
논문 발표회									——

#### 3.1 Overview

[표 1]은 이번 달부터 논문을 발표할 11월까지의 대략적인 계획을 설명한 표이다. 우선 다가오는 23일 까지 서약서 및 제안서를 작성 완료한다. 이후 논문 작성을 위한 공공 데이터를 수집하고 이를 분석할 도구인 분석기법을 연구 할 것이다.

이때 좀 더 의미 있는 데이터를 도출해 내기 위해 이미 과거부터 경제적으로 성장해 있던 지역인 특별시, 광역시 및 기타 우리나라 주요 도시들은 최대한 배제하고 과거와 현재의 경제성장정도가 많이 차이나는 중소 시/군/구를 중심으로 데이터를 조사할 계획이다. 사전에 검색해 본 바에 따르면 특별시, 광역시 및

주요 도시에서의 데이터가 그렇지 않은 곳들의 데이터에 비해 훨씬 더 다양하고 방대하다는 것을 알 수 있었다. 이는 빅 데이터의 중요성이 대두된 지 그리 오래되지 않았고 만약 존재한다 하더라도 위에서 언급한 주요 도시 위주의 자료일 것이라는 점을 감안하면 당연한 결과일지 모른다. 하지만 앞서 말한 것처럼 이미 성장된 지역의 성장정도를 측정하여 비만율과 상관관계를 찾는 것은 그 관계가 두드러지지 않을뿐더러 큰 의미가 있다고 생각하지 않기 때문에 기존에 계획했던 대로 이전에는 그렇지 않았지만 현재 어느 정도 발전된 중소도시를 위주로 데이터를 분석하기로 했다.

이후 연구한 분석기법들을 이용하여 데이터를 본격적으로 분석하고 의미 있는 결과 데이터를 도출해 낼 것이다. 분석된 데이터에 따라 논문을 작성하고 중간보고서와 최종보고서를 통해 새로 추가된 내용이 있거나 부족한 부분이 있으면 수정함으로써 논문작성을 마무리 할 것이다.

## 4. 기대효과 및 개선방향

### 4.1 기대 효과

본 연구는 경제적인 부분을 포함하여 다양한 사회적 요인의 변화가 어떻게 비만율의 변화에 영향을 미쳐왔는지에 대해 연구를 진행한다. 위에서 언급했다시피 비만은 현재 국제사회의 심각한 문제들 중 하나이다. 이 둘 사이의 뚜렷한 상관관계가 데이터 분석을 통해 도출이 된다면 비만문제를 사회 여러 분야에서 다각적으로 접근해야 할 문제임을 인식시키는데 도움을 줄 수 있고 이를 통해 비만문제 해결 대책을 수립하는데 도움을 줄 수 있을 것이다. 또한 비만문제 해결을 통해 삶의 질을 높이는데 기여를 할 수 있을 것이라 생각된다.

### 4.2 개선 방향

선행연구 부분에서 말한 것처럼 우리나라 주요 도시들에 비해 중소도시들의 공공 데이터는 상대적으로 많이 부족하다. 특히 도시성장의 중요한 지표로 볼 수 있는 지역내총생산(GRDP), 행정구역별 1인당 개인소득 등의 중소도시들의 중요한 경제지표로 볼 수 있는 데이터들은 통계청 및 다른 공공기관에서 제공하지 않는다. 어떻게 보면 가장 중요한 지표로 여겨질 수 있는 부분인데 상당히 아쉽다. 따라서 앞서 찾아보지 못한 도시의 성장을 측정할 수 있는 사회적, 경제적 척도들을 새롭게 찾아내는 것이 중요하다고 생각한다.

## 5. 기타

해당 없음

## 6. 참고문헌

- [1] 이경호, "21 세기 한국사회 전망/[복지분야] 21 세기 사회에 대비한 보건복지분야의 주요 정책과제." 한국행정연구 9,1 (2000)
- [2] 이명해, "비만의 원인과 평가에 관한 고찰." 부산여자전문대학 논문집 17 (1995)
- [3] Prevalence of overweight <http://www.oecd.org/health/obesity-update.htm/>
- [4] Obesity in America  
<http://www.centralohiobariatrics.com/obesity-in-america/>
- [5] 박성복, "지역발전도 및 지역불평등도의 측정." 한국행정학보 31.3 (1997)
- [6] Tom M. Mitchell, Machine Learning (McGraw-Hill International Editions Computer Science Series) 1st Edition (McGraw-Hill,1997)
- [7] Huang, Hsin-Hsiung, Su-Yun Huang, "Nonlinear Regression Analysis." International Encyclopedia of Education (2010)
- [8] Nazeer, KA Abdul, M. P. Sebastian, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm." Proceedings of the world congress on engineering Vol 1 (2009)