

원산지 위반 • 단속 패턴 종합 분석 및  
SI를 활용한 농산물 원산지 단속추천 서비스



빅리더 인스티튜트

# 제 출 문

국립농산물품질관리원장 귀하

본 보고서를 「원산지 위반·단속 패턴 종합분석 및 AI를 활용한  
농산물 원산지 단속추천 서비스」에 관한 연구용역의 최종보고서로  
제출합니다.

2019년 11월

연구기관: 빅리더 인스티튜트  
연 구 진: 전종식 (연구책임자)  
정진교 (공동연구자)  
박준하 (연구원)  
최재원 (연구원)  
안준영 (연구원)  
염재영 (연구원)  
박성희 (연구원)  
신재원 (연구원)  
박상욱 (연구원)  
장서진 (연구원)  
한인용 (연구원)  
서강원 (연구원)  
협력 김지성(외부연구원)

## 목 차

### I 서론

1 연구배경 및 필요성 .....	1
가. 빅데이터 산업 및 스마트 AI 농업 행정 .....	1
나. 원산지 .....	2
다. 원산지 표시제도 도입 배경 및 실태 .....	2
라. 원산지 표시 관련 법령 .....	4
마. 원산지 조사 실태 및 연구 필요성 .....	5
바. 원산지 표시제도 현황 .....	6
사. 원산지 표시 조기경보 서비스 .....	10
2 연구 개요 .....	11
가. 연구 목적 .....	11
나. 연구 배경 .....	11
다. 연구 진행 방향 .....	11
3 연구 범위와 연구 방법 .....	13
가. 연구 범위 .....	13
나. 연구 방법 .....	14

### II 인공의 지능 추천서비스 사례연구

1 국내사례 .....	15
2 해외사례 .....	16

### III 원산지 조사 및 위반 단속 분석

가. 전국 현황 .....	22
나. 대도시별 단속 현황 .....	25
다. 연도별 단속 비율 .....	27
라. 연도별 단속 건수 .....	28
마. 월별 단속 비율 .....	36
바. 요일별 단속 현황 .....	44
사. 단속 현황 .....	48

아. 5대 품목별 단속 현황 .....	63
자. 위반 금액별 단속 현황 .....	70
차. 업체별 단속 현황 .....	71
카. 제위반 경향성 분석 .....	80

#### IV 활용데이터 설명 및 현황

1 업체 및 지역 .....	87
가. 업체추천 데이터 .....	87
나. 지역추천 데이터 .....	90
2 외부요인 분석 - 뉴스분석 .....	93

#### V 알고리즘 설명

1 머신러닝을 활용한 단속업체 추천 .....	96
가. 모형 개요 .....	96
나. 머신러닝 모형 종류 .....	96
다. 업체추천 분석 활용변수 .....	103
라. 적용과정 .....	105
2 딥러닝(CNN)을 활용한 단속지역 추천 .....	107
가. 모형 개요 .....	107
나. CNN 구성 .....	108
다. 지역추천 분석 활용변수 .....	111
라. CNN 모형 적용과정 .....	113

#### VI 분석결과

1 단속 업체추천 .....	117
가. 머신러닝을 활용한 업체추천 결과 .....	117
나. 머신러닝을 활용한 업체추천 결과 요약표 .....	132
2 단속 지역추천 .....	134
가. CNN을 활용한 지역추천 결과 .....	134

### II

#### VII 원산지 표시 단속추천 시스템 활용 및 구축방안

1 단속추천 시스템 활용방안 .....	138
가. 업체·지역추천 시스템 활용 .....	138
나. 맵 커스터마이징 기능 .....	140
다. 실시간으로 활용 가능한 대시보드 기능 .....	141
2 단속추천 시스템 구축방안 .....	142
가. 현황 .....	142
나. 단속추천 시스템 개발 개요 .....	143
다. 단속추천 시스템 배포 및 관리 .....	146
라. 사용 프로그램 .....	148

#### VIII 요약 및 결론

1 요약 .....	153
2 결론 .....	155
3 제언 .....	157

### III

## 표 목 차

<표 1-1> 2019년 11월 농식품 원산지 표시 조기경보 서비스 .....	10
<표 2-1> 해외사례 .....	20
<표 3-1> 단속 업주별 성별 .....	22
<표 3-2> 단속 업주 나이 .....	23
<표 3-3> 지역별 단속 업주 나이 .....	24
<표 3-4> 대도시별 단속 현황 .....	25
<표 3-5> 연도별 단속 비율 .....	27
<표 3-6> 연도별 단속 건수 .....	28
<표 3-7> 전국 연도별 단속 내역 .....	29
<표 3-8> 광주광역시 연도별 단속 내역 .....	30
<표 3-9> 대구광역시 연도별 단속 내역 .....	31
<표 3-10> 대전광역시 연도별 단속 내역 .....	32
<표 3-11> 부산광역시 연도별 단속 내역 .....	33
<표 3-12> 서울특별시 연도별 단속 내역 .....	34
<표 3-13> 인천광역시 연도별 단속 내역 .....	35
<표 3-14> 월별 단속 비율 .....	36
<표 3-15> 전국 월별 단속 내역 .....	37
<표 3-16> 광주광역시 월별 단속 내역 .....	38
<표 3-17> 대구광역시 월별 단속 내역 .....	39
<표 3-18> 대전광역시 월별 단속 내역 .....	40
<표 3-19> 부산광역시 월별 단속 내역 .....	41
<표 3-20> 서울특별시 월별 단속 내역 .....	42
<표 3-21> 인천광역시 월별 단속 내역 .....	43
<표 3-22> 전국 요일별 단속 현황 .....	44
<표 3-23> 정기단속 월별 주요 내용 .....	50
<표 3-24> 전국 5대 품목별 단속 현황 .....	63
<표 3-25> 광주광역시 5대 품목별 단속 현황 .....	64
<표 3-26> 대구광역시 5대 품목별 단속 현황 .....	65
<표 3-27> 대전광역시 5대 품목별 단속 현황 .....	66
<표 3-28> 부산광역시 5대 품목별 단속 현황 .....	67
<표 3-29> 서울특별시 5대 품목별 단속 현황 .....	68

## IV

<표 3-30> 인천광역시 5대 품목별 단속 현황 .....	69
<표 3-31> 지역별 위반 금액 비중 .....	70
<표 3-32> 전국 업체별 단속률 .....	72
<표 3-33> 전국 업체 분포 .....	73
<표 3-34> 서울 업체 분포 .....	74
<표 3-35> 인천 업체 분포 .....	75
<표 3-36> 부산 업체 분포 .....	76
<표 3-37> 대전 업체 분포 .....	77
<표 3-38> 대구 업체 분포 .....	78
<표 3-39> 광주 업체 분포 .....	79
<표 3-40> 지역별 재적발률 .....	80
<표 3-41> 업주 연령에 따른 제위반율 .....	81
<표 3-42> 위반 물량별 재적발률 .....	83
<표 3-43> 전국 위반 금액별 제위반율 .....	85
<표 4-1> 활용 데이터 목록(업체추천) .....	87
<표 4-2> 업체추천 분석 데이터 출처 및 설명 .....	88
<표 4-3> 업체추천 분석 데이터 출처 및 설명(계속) .....	89
<표 4-4> 활용 데이터 목록(지역추천) .....	90
<표 4-5> 지역추천 분석 데이터 출처 및 설명 .....	91
<표 4-6> 검색어 별 수집기사 건수 .....	94
<표 4-7> ARIMA 분석결과 .....	95
<표 5-1> 활용변수표 .....	104
<표 5-2> 활용변수표 .....	112
<표 6-1> 서울특별시 업체추천 모형별 f1스코어 .....	117
<표 6-2> 분류결과표(Confusion_Matrix) .....	118
<표 6-3> 분류결과표의 f1스코어 .....	119
<표 6-4> 대전광역시 업체추천 모형별 f1스코어 .....	120
<표 6-5> 분류결과표(Confusion_Matrix) .....	120
<표 6-6> 분류결과표의 f1스코어 .....	121
<표 6-7> 광주광역시 업체추천 모형별 f1스코어 .....	122
<표 6-8> 분류결과표(Confusion_Matrix) .....	122
<표 6-9> 분류결과표의 f1스코어 .....	123
<표 6-10> 인천광역시 업체추천 모형별 f1스코어 .....	124

## V

<표 6-11> 분류결과표(Confusion_Matrix) .....	125
<표 6-12> 분류결과표의 f1스코어 .....	125
<표 6-13> 부산광역시 업체추천 모형별 f1스코어 .....	126
<표 6-14> 분류결과표(Confusion_Matrix) .....	127
<표 6-15> 분류결과표의 f1스코어 .....	127
<표 6-16> 대구광역시 업체추천 모형별 f1스코어 .....	128
<표 6-17> 분류결과표(Confusion_Matrix) .....	128
<표 6-18> 분류결과표의 f1스코어 .....	129
<표 6-19> 주요 6대 도시 업체추천 모형별 f1스코어 .....	130
<표 6-20> 분류결과표(Confusion_Matrix) .....	131
<표 6-21> 분류결과표의 f1스코어 .....	131
<표 6-22> 분류결과표의 f1스코어 .....	132
<표 6-23> 변수 판단 개수 .....	134
<표 6-24> 분류결과표(Confusion_Matrix) .....	135
<표 6-25> 분류결과표의 f1스코어 .....	135
<표 7-1> 단계별 시스템 구축 계획(예시) .....	145
<표 7-2> 국가에서 제공하는 클라우드 서비스 예시 .....	146
<표 7-3> 원산지 표시 단속추천 시스템 구축의 필요 프로그램 .....	152

## VI

### 그림목차

<그림 1-1> 원산지 표시 예시 .....	6
<그림 1-2> 농관원 홈페이지 공표 .....	7
<그림 1-3> 육안으로 구별이 어려운 국내산(왼쪽)과 덴마크산 삼겹살(오른쪽) .....	8
<그림 1-4> 연구 진행 방향 .....	12
<그림 1-5> 전체 품목 대비 상위 대비 상위품목 비율 .....	13
<그림 1-6> 연구개요 .....	14
<그림 3-1> 단속 업주별 성별 .....	22
<그림 3-2> 단속 업주 나이 .....	23
<그림 3-3> 지역별 단속 업주 나이 .....	24
<그림 3-4> 단속률 .....	26
<그림 3-5> 주요 도시 단속률 .....	26
<그림 3-6> 연도별 단속 비율 .....	27
<그림 3-7> 전국 연도별 단속 내역 .....	29
<그림 3-8> 광주광역시 연도별 단속 내역 .....	30
<그림 3-9> 대구광역시 연도별 단속 내역 .....	31
<그림 3-10> 대전광역시 연도별 단속 내역 .....	32
<그림 3-11> 부산광역시 연도별 단속 내역 .....	33
<그림 3-12> 서울특별시 연도별 단속 내역 .....	34
<그림 3-13> 인천광역시 연도별 단속 내역 .....	35
<그림 3-14> 월별 단속 비율 .....	36
<그림 3-15> 전국 월별 단속 내역 .....	37
<그림 3-16> 광주광역시 월별 단속 내역 .....	38
<그림 3-17> 대구광역시 월별 단속 내역 .....	39
<그림 3-18> 대전광역시 월별 단속 내역 .....	40
<그림 3-19> 부산광역시 월별 단속 내역 .....	41
<그림 3-20> 서울특별시 월별 단속 내역 .....	42
<그림 3-21> 인천광역시 월별 단속 내역 .....	43
<그림 3-22> 전국 요일별 단속 현황 .....	44
<그림 3-23> 서울특별시 요일별 단속 현황 .....	45
<그림 3-24> 부산광역시 요일별 단속 현황 .....	45
<그림 3-25> 인천광역시 요일별 단속 현황 .....	46

## VII

<그림 3-26> 대구광역시 요일별 단속현황 .....	46
<그림 3-27> 대전광역시 요일별 단속현황 .....	47
<그림 3-28> 광주광역시 요일별 단속현황 .....	47
<그림 3-29> 1월 정기 단속 현황 .....	53
<그림 3-30> 2월 정기 단속 현황 .....	54
<그림 3-31> 3월 정기 단속 현황 .....	55
<그림 3-32> 4월 정기 단속 현황 .....	56
<그림 3-33> 5월 정기 단속 현황 .....	57
<그림 3-34> 7월 정기 단속 현황 .....	58
<그림 3-35> 8월 정기 단속 현황 .....	59
<그림 3-36> 9월 정기 단속 현황 .....	60
<그림 3-37> 11월 정기 단속 현황 .....	61
<그림 3-38> 12월 정기 단속 현황 .....	62
<그림 3-39> 전국 5대 품목별 단속 현황 .....	63
<그림 3-40> 광주광역시 5대 품목별 단속 현황 .....	64
<그림 3-41> 대구광역시 5대 품목별 단속 현황 .....	65
<그림 3-42> 대전광역시 5대 품목별 단속 현황 .....	66
<그림 3-43> 부산광역시 5대 품목별 단속 현황 .....	67
<그림 3-44> 서울특별시 5대 품목별 단속 현황 .....	68
<그림 3-45> 인천광역시 5대 품목별 단속 현황 .....	69
<그림 3-46> 지역별 위반 금액 비중 .....	70
<그림 3-47> 업태 축소 현황 .....	71
<그림 3-48> 전국 업태별 단속률 .....	72
<그림 3-49> 전국 업태 분포 .....	73
<그림 3-50> 서울 업태 분포 .....	74
<그림 3-51> 인천 업태 분포 .....	75
<그림 3-52> 부산 업태 분포 .....	76
<그림 3-53> 대전 업태 분포 .....	77
<그림 3-54> 대구 업태 분포 .....	78
<그림 3-55> 광주 업태 분포 .....	79
<그림 3-56> 도시별 제위반율 .....	80
<그림 3-57> 업주 연령별 제위반율 .....	82
<그림 3-58> 전국 위반 물량별 제위반율 .....	84

## VIII

<그림 3-59> 전국 위반 금액별 제위반율 .....	86
<그림 4-1> 뉴스 분석 진행도 .....	93
<그림 4-2> 감성점수 도출 예시 .....	94
<그림 4-3> 월별 뉴스 점수 증감을 품목별 단속 건수 증감을 비교 .....	95
<그림 5-1> KNN의 분류 방식 .....	97
<그림 5-2> 의사결정나무 분류 과정 그래프 .....	99
<그림 5-3> 다층 퍼셉트론 예측 과정 .....	101
<그림 5-4> SVM 분류 방식 .....	102
<그림 5-5> 전처리하기 전 데이터 .....	105
<그림 5-6> 3이 적혀있는 이미지를 CNN 계층을 통해 분류 .....	107
<그림 5-7> 33개의 채널로 만들어진 컬러사진 .....	108
<그림 5-8> 2차원 입력 데이터를 1개의 필터로 합성곱 계산을 수행 .....	109
<그림 5-9> Stride가 1로 필터를 입력 데이터에 이동하는 과정 .....	109
<그림 5-10> 데이터의 외곽에 패딩을 사용하여 데이터 크기를 유지 .....	110
<그림 5-11> MaxPooling과 AveragePooling의 예시 .....	110
<그림 5-12> 전처리 전 데이터 .....	113
<그림 5-13> MinMaxScaler사용 전과 후 .....	114
<그림 5-14> CNN 2층 구상도 .....	115
<그림 6-1> Convolution Layer 훈련 그래프 .....	134
<그림 6-2> 왼쪽 그래프: 18년 실제값, 오른쪽 그래프: 18년 예측값 .....	137
<그림 7-1> 시안 프로그램 업체추천 현황 (영등포구) .....	138
<그림 7-2> 시안 프로그램추천 업체 정보 (영등포구) .....	139
<그림 7-3> 시안 프로그램 지역추천 현황 (영등포구) .....	139
<그림 7-4> 시안 프로그램 클러스터링 기능 .....	140
<그림 7-5> 시안 프로그램 대시보드 - 서울시 구별 추천업체 현황 .....	141
<그림 7-6> 시안 프로그램 대시보드 - 서울시 추천업체 .....	141
<그림 7-7> 세 가지의 프리 티어 오피 유형 .....	147
<그림 7-8> QGIS와 ArcGIS .....	148
<그림 7-9> 텐서플로우 및 사이킷런 .....	149
<그림 7-10> Folium .....	149
<그림 7-11> Django .....	150
<그림 7-12> KT G-Cloud .....	150
<그림 7-13> PostgreSQL .....	150

## IX

X

## I 서론

### 1 연구배경 및 필요성

#### 가. 빅데이터 산업 및 스마트 AI 농업 행정

- 사회혁신(Social Innovation)은 기존의 방식으로 해결하기 어렵거나 새롭게 발생하는 사회문제들을 '새로운 방법 (New ideas)'으로 해결하는 것을 의미함. 이는 단순히 창조적이고 혁신적인 개인의 아이디어를 적용해 보는 것이 아니라 사회의 전반적인 시스템을 변화시켜 공공선을 실현하는 것을 의미하는 것으로 이러한 사회혁신은 사회문제 해결을 위한 글로벌 차원의 새로운 대안임
- 그동안 빅데이터란 무엇인지 과연 경제적 효과가 있는지 등 빅데이터의 버블 논란이 한동안 지속되었지만 2000년대 들어 정보기술(IT)이 발전하면서 데이터가 기업의 이익으로 연결되면서 데이터 과학에 대한 관심이 쏠림. 특히 넷플릭스와 아마존이 고객의 구매 패턴을 분석한 뒤 맞춤형 추천서비스로 성공을 하자 글로벌 기업들이 데이터 사이언스에 주목하기 시작함
- 전 세계적으로 빅데이터의 실체에 대한 논란에서 벗어나 데이터의 실질적 가치를 발굴하고 비즈니스로 연결시켜 데이터와 AI가 하나의 산업으로서 자리매김하고 있는 상황임
- 현재 데이터 산업은 진화 단계에 진입하면서 기업의 데이터 수요는 급증하고 데이터 수익화를 위한 다양한 비즈니스가 활성화되고 있음. Data Science, AI, 빅데이터의 도입은 미디어/커뮤니케이션 금융, 공공, 서비스 산업을 넘어 비용 절감 효과를 누릴 수 있는 전자, 소재 산업 등 제조업 전반으로 확산 중이며 또한 공공 부문으로 데이터 산업이 들어옴으로써 실질적으로 적용하려는 스마트 AI 행정의 필요성이 강하게 제기되고 있는 상황임
- 국민 먹거리 부분에서 수입의 의존도가 높은 우리로서는 농업 부분에 있어서 빅데이터 및 AI의 적용을 통해 스마트 AI 농업 행정의 적

용과 활용이 그 어느 때보다 더 필요한 상황이라고 할 수 있음

## 나. 원산지

- 원산지란 농산물이 생산·채취된 국가 또는 지역, 국제적 거래에서는 일반적으로 그 물품이 생산된 정치적 실체를 지닌 국가를 가리키고 국내적으로는 지역 또는 지방을 의미함. 가공·생산공정 또는 재배 등의 과정을 거치지 않고 단순히 그 국가를 통하여 거래되었음을 의미하는 경유국, 적출국, 수출국과는 다른 개념임
- 수입 농산물의 원산지 표시는 국가 간의 수출입 상거래질서 확립 및 생산자, 특히 농업 생산자와 소비자의 권익보호 등의 원산지 표시의 기능과 역할을 함. 이러한 표시규정을 위반하게 될 경우 그 영향은 크고 중대하게 나타난다고 할 수 있음
- 원산지 표시제도는 국제규범에서 허용하고 있는 제도로서 미국, EU, 일본 등 대부분의 국가가 원산지 표시 제도를 운영하고 있음

## 다. 원산지 표시제도 도입 배경 및 실태

- 원산지 표시제도의 목적은 국내에서 유통되는 농산물 및 그 가공품과 음식점에서 판매되는 쌀, 배추김치, 콩, 축산물 등에 대한 원산지 표시 관리로 소비자의 알 권리와 선택권을 보장하고 유통질서를 확립하여 생산자와 소비자 보호하기 위함임<sup>1)</sup>
- 수입·판매되는 농축산물이 동일 또는 유사한 형질을 가지고 있더라도 소비자가 판단하는 가치가 원산지에 따라 다르고 국민들의 선호도가 판이하게 달라지므로 원산지의 정확한 표시는 국민의 권익보호 및 국내 농축산업 보호에 중요한 의미를 가짐. 그러나 그 취지에도 불구하고 매년 원산지 표시 위반이 적발되고 있어 소비자 피해가 지속되고 있는 상황임

1) 양성법, 농식품 원산지표시 제도 개선방안 연구, 단국대학교(2018), 1쪽

## 2

- 국내 유통과정에서 수입품의 원산지를 미표시하거나 국산이나 외국산으로 속여서 판매하는 등 불법적인 원산지 표시와 그 피해 사례가 자주 언론에 보도되는 등 사회문제로 제기되고 있음  
- 연평균 적발금액: 약 136억원
- 원산지 표시제도는 1991년 대외무역법에 의해 본격적으로 도입되었으며 관세청이 수입통관 단계에서만 불법 원산지 표시 물품의 통과 차단에 관여해 왔으나 2007년 4월 이후부터 국내유통 단계 수입 물품의 원산지 표시 위반에 대해서도 단속을 실시하고 있음<sup>1)</sup>
- 먹거리에 관한 불량·유해 수입 물품의 경우 원산지 식별이 용이하지 않은 점을 악용하여 국내·외 물품의 혼합, 지역특산물 위장 등 원산지 둔갑 수법은 갈수록 지능화되고 있는 추세임

1) 정재완, 시중유통 수입물품 원산지표시 단속체계의 효율화 방안, 한남대학교 산업협력단(2009), 181쪽

## 3



라. 원산지 표시 관련 법령<sup>1)</sup>

- 원산지 표시대상
  - 농수산물[농수산물의 원산지 표시에 관한 법(이하 법이라 한다) 제5조, 농수산물의 원산지 표시에 관한 법률 시행령(이하 시행령이라 한다) 제3조]
  - 농수산물 가공품 및 그 원료(법 제5조, 시행령 제3조)
  - 대외무역법 제33조에 따라 산업통상자원부장관이 공고한 수입 농수산물 또는 그 가공품(시행령 제3조)
  - 식품위생법 시행령 중 휴게음식점영업, 일반음식점영업, 위탁급식영업, 집단급식소에서 농수산물이나 그 가공품을 조리하여 판매·제공하는 경우에 그 농수산물이나 그 가공품의 원료(법 제5조)
- 원산지 표시를 하여야 할 자(법 제5조, 시행령 제4조)
  - 농수산물 또는 그 가공품을 수입하는 자
  - 생산·가공하여 출하하거나 판매(통신판매를 포함)하는 자
    - 통신판매: 「전자상거래 등에서의 소비자보호에 관한 법률」 제12조에 따라 신고한 통신판매업자의 판매(전단지를 이용한 판매 제외)또는 같은 법 제20조 제2항에 따른 통신판매중개업자가 운영하는 사이버몰을 이용한 판매를 말함(법 제2조, 시행령 제2조)
  - 판매할 목적으로 보관·진열하는 자
  - 식품접객업 및 집단급식소 중 휴게음식점영업, 일반음식점영업, 위탁급식영업소, 집단급식소를 설치·운영하는 자
- 거짓 표시 등의 금지(법 제6조)
  - 모든 영업자
    - 원산지 표시를 거짓으로 하거나 이를 혼동하게 할 우려가 있는 표시를 하는 행위
    - 원산지 표시를 혼동하게 할 목적으로 그 표시를 손상·변경하는 행위
    - 원산지를 위장하여 판매하거나, 원산지 표시를 한 농수산물이나 그 가공품에 다른 농수산물이나 가공품을 혼합하여 판매하거나 판매할 목적으로 보관이나 진열하는 행위

1) 양성법, 농식품 원산지표시 제도 개선방안 연구, 단국대학교(2018년), 1-2쪽

- 농수산물이나 그 가공품을 조리하여 판매·제공하는 자
- 원산지 표시를 거짓으로 하거나 이를 혼동하게 할 우려가 있는 표시를 하는 행위
- 원산지를 위장하여 조리·판매·제공하거나, 조리하여 판매·제공할 목적으로 농수산물이나 그 가공품의 원산지 표시를 손상·변경하여 보관·진열하는 행위
- 원산지 표시를 한 농수산물이나 그 가공품에 원산지가 다른 동일 농수산물이나 그 가공품을 혼합하여 조리·판매·제공하는 행위

마. 원산지 조사 실태 및 연구 필요성<sup>1)</sup>

- 원산지 표시 위반 및 이에 대한 단속에 대한 여론과 언론의 관심은 항상 높은 편이며 생산자단체와 소비자단체가 강력한 단속을 지속적으로 요구하고 있는 상황임
- 수입 농산물의 원산지 표시 위반에 대한 단속은 근래 새롭게 부여된 중요한 업무영역의 하나이자 일반 국민의 피부에 직접 닿는 행정으로 새로운 가치를 창출시킬 수 있는 의미 있는 업무 분야로 부상하고 있음
- 이에 국립농산물품질관리원(이하 농관원)은 각 지원에 기동단속반, 시군 사무소에 일반단속반을 편성하고 원산지 표시 조사를 시행 중이나 많은 어려움을 겪고 있음
  - 조사 대상 품목 대비 턱없이 부족한 조사 인력
    - 조사 대상 대비 조사 인원 약 3.7%
  - 노점상과 전통시장에 대한 조사의 한계
  - 과태료 미납 및 체납자에 대한 강제 징수문제
  - 개개인의 경험과 관행에 의존한 진행
- 따라서 그 원인에 대한 분석과 해결을 위한 대안을 심층 연구하여 효율적이고 효과적인 원산지 표시 위반 조사방안 마련이 필요하므로

1) 국립농산물품질관리원, 2015 농산물품질관리연보, 농림수산물부(2016), 117쪽

빅데이터 정보에 대한 종합분석을 통하여 다양한 위반·단속 패턴을 밝혀내고 이를 토대로 향후 「원산지 단속추천 시스템」 개발을 위한 사전 기반 구축을 위해 본 연구를 진행함

## 바. 원산지 표시제도 현황

- 일본 후쿠시마 원전사고 이후 원산지 표시에 관한 소비자 관심도가 높아지고 있음(2018년 농관원 농식품 부정유통 신고 처리 건수: 1,411건)
- 지난 일본 원전사고 이후로 일본 등 해외 수산물이 원산지를 국내로 거짓 표시하여 유통된 사실이 국립수산물품질관리원의 단속 결과 대거 확인되었음<sup>1)</sup>
- 이에 따라 소비자의 수입 농식품 안전에 대한 불안감이 고조되고 있으며 동시에 원산지 표시 제도에 대한 관심은 증가하고 있음
- 원산지 거짓표시 적발 등에 따른 사건 이후 소비자들은 원산지 표시 제도를 긍정적으로 평가하고 있으며, 표시제도 확대를 요구하고 있음

<그림 1-1> 원산지 표시 예시



1) 최훈길, <국산으로 둔갑한 日수산물...농·수협도 원산지 '허위 표시'>, 《이데일리》, 2019.04.19.(<https://www.edaily.co.kr/news/read?newsId=01167686622457104&mediaCodeN=257&OutLnkChk=Y>), 검색일: 2019.07.09

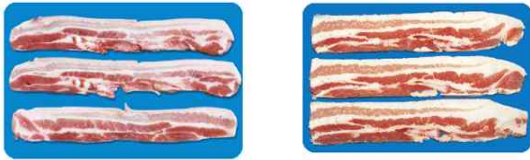
- 시민들이 자신이 이용하는 음식점, 정육점이 원산지 표시제도를 잘 이행했는지에 대해 확인하는 것이 어렵고 이런 정보에 대한 접근이 어려워 자발적 견제가 어렵기는 하나 최근 관심도는 높아지고 있음
- 예를 들어 2012.1.26.부터 법 개정으로 원산지 위반정보에 대한 인터넷 공표범위가 확대됨에 따라 원산지 위반정보를 농관원의 홈페이지에서 쉽게 확인할 수 있게 되었지만, 일반인이 정보를 확인하고자 하는 의지가 낮은 실정임

<그림 1-2> 농관원 홈페이지 공표



- 수입 축산물의 경우 실질적으로 소비자가 육안으로 원산지를 구별하는 것은 현실적으로 불가능하기 때문에 쇠고기나 돼지고기의 경우 판매자가 원산지를 속이는 경우 소비자는 속고 살 수 밖에 없는 상황임

<그림 1-3> 육안으로 구별이 어려운 국내산(왼쪽)과 덴마크산 삼겹살(오른쪽)



- 농관원의 제한된 단속 인원 및 업무 여건 때문에 모든 지역을 충분히 관리하는 것이 힘든 상황임
  - 원산지 표시 대상 업체는 전국에 150여 만개에 달함
  - 농관원은 소비자단체와 농민단체 관계자 등 1만2천여명을 명예감시원으로 위촉하고 수시로 원산지 표시 위반행위를 단속하고 있으나 농관원의 전담 인력 270여명이 전국 150여만개에 달하는 대상 업체를 단속하기엔 역부족인 상황임
- 농관원은 단속과 처벌이라는 사후 대책뿐만 아니라 위반 자체를 예방하는 사전 대책 또한 병행되어야 함
  - 단속의 목적은 물론 그 위반상태를 제거하는 것에도 있지만, 위반상태 제거만을 목적으로 하는 무작위적인 단속은 그 위반상태를 제거함에 있어서 일회적인 가능성이 높고 이에 따라서 제재 위주의 단속보다는 예방 위주의 단속이 그 효용성을 가진다고 할 것임. 이런 측면에서 전통시장 상인 교육 등과 같은 여러 가지 정책이 실행되고 있기는 하지만 부족한 부분이 많은 것이 사실이며 노점상과 전통시장 상인에 대해 지속적으로 원산지를 표시하도록 하게 하는 유인책 또는 방법이 강구되어야 함

## 8

- 농관원의 단속 및 위반 신고 처리 절차
  - 부정유통신고 담당(야간에는 당직근무자가 접수하여 익일 업무 담당자에게 인계인수)자가 접수
  - 인터넷 신고건은 본원 원산지관리과에서 매일 2회(9시, 13시)이상 검색하여 해당 지원(또는 사무소)에 통지함
- 담당 공무원이 현장에 출동하여 업주 임회하여 현물 확인과 장부 등을 조사하고 현물 사진촬영, 시료채취 등 위반행위 증거 확보 후 확인서를 정구함
  - 거짓표시 행위자는 형사 입건하여 수사 후 검찰에 송치하고 미표시 행위자는 과태료부과 조치함
  - 신고자에게 조치결과를 알려주고 형사처벌 또는 과태료부과 처분이 종결되면 기준에 따라 원산지 위반 신고포상금 지급함
- 농관원의 위반 시 처벌내용은 다음과 같음<sup>1)</sup>
  - 형사처벌: 7년 이하의 징역 또는 1억원 이하의 벌금(상습위반자: 1년 이상 10년 이하의 징역 또는 500만원 이상 1억 5천만원 이하의 벌금 부과)
    - \* 행위자: 원산지 거짓표시, 혼동 우려 표시 행위, 원산지를 혼동하게 할 목적으로 그 표시를 손상 또는 변경함. 원산지를 위장·혼합하여 판매하거나 판매할 목적으로 보관 또는 전열함
  - 거짓표시자 과징금 부과: 위반금액의 5배 이하(최고 3억원)부과
    - \* 행위자: 2년간 2회 이상 거짓표시로 적발
  - 원산지 표시제도 교육 이수명령: 3개월 이내 의무교육 이수
    - \* 행위자: 미표시 2년 이내에 2회 이상 또는 거짓표시 위반
  - 과태료부과: 5만원 이상 1천만원 이하 음식점
    - \* 행위자: 농산물 및 가공품 원산지 미표시

1) 오원석, 무역환경변화에 대응한 효과적인 원산지제도 개선에 관한 연구, 한국무역상무학회(2008), 149쪽

## 사. 원산지 표시 조기경보 서비스

- 농관원에서는 매월 <표 1-1>과 같은 형식으로 농식품 원산지 표시 조기경보 서비스를 제공하고 있음. 이는 농관원·지자체 등 단속 공무원에게 원산지 표시 위반 개연성이 높은 농식품 품목과 식별정보를 제공하여 조사단속의 효율성을 높이기 위한 것임

<표 1-1> 2019년 11월 농식품 원산지 표시 조기경보 서비스

단 계	품 목
경 보 (30% 이상)	녹두, 콩, 낙화생(탈각), 들깨, 감귤, 배춧속채소(꽃양배추, 구경양배추, 케일), 생강, 도라지, 마늘(냉동), 장미(절화), 카네이션, 부자, 패모, 두충, 산사자, 사인, 연자육, 오리고기, 녹용
주 의 (10% 이상~30% 미만)	팥, 참깨, 시금치류, 멜론, 켄(건조/참취, 개미취, 곰취, 미역취), 메주, 녹차, 면양고기, 돼지고기, 닭고기, 칠면조, 장(소·돼지의 것), 목이버섯(건조), 대추(건조), 두부
관 심 (0% ~10% 미만)	강낭콩, 낙화생(미탈각), 무화과(신선), 국화(절화), 안개초(절화), 감초, 구기자, 포도, 밤, 양파, 고추류(냉동), 쇠고기

- 전년 동기 대비 수입물량 증가율과 국산/외국산 가격 차이에 따라 경보·주의·관심 단계로 구분하고 해당 월에 경보

10

## 2 연구 개요

### 가. 연구 목적

- 원산지 표시 조사 효율성의 극대화를 위해 데이터 과학에 입각한 효율적인 수사방안 도출을 통한 계획적이고 체계적인 단속으로의 패러다임 전환
- 다양한 원산지 관련 빅데이터 정보에 대한 종합분석을 통하여 위반·단속 패턴을 찾아내어 향후 「원산지 단속추진 시스템」 개발을 위한 사전 기반 구축함

### 나. 연구 배경

- 농관원의 원산지 단속 약 10년간(2009-2018)의 축적된 빅데이터를 종합적으로 분석하여 효율적인 단속·수사 방안 도출이 필요함
- 기존의 관행적 또는 신고 위주의 원산지 단속에서 과학적 데이터 분석 기반의 계획적이고 체계적인 단속으로의 패러다임 전환 요구됨
- 궁극적으로 지도기반 위에서 신속·정확하게 단속 대상을 추천받아 효율적으로 조사하는 선제적 사전 단속방법 개발이 필요함

### 다. 연구 진행 방향

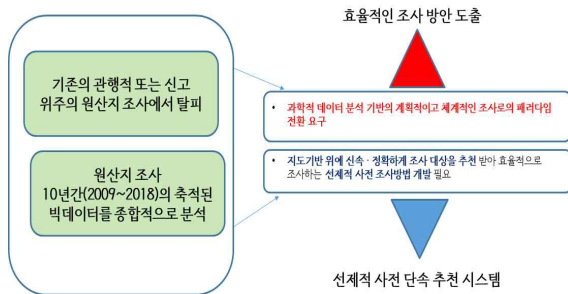
- 농관원 및 지자체 등에서 보유하고 있는 2009년에서 2018년, 10개년도 조사·단속 데이터와 공공데이터 등 빅데이터의 분석을 통한 위반행위 패턴 파악 및 선제적 단속방법 개발
- 조사정보의 시계열 자료와 지역 데이터를 기반으로 지역 정보 분석, 통계분석, 기계학습 등 심층분석을 실시함
- 범법자의 유형 및 위반이 발생하는 지역적 요인 등을 도출
- 이를 기반으로 시기, 지역, 품목 등을 고려한 원산지 조사추진 및 선

11

제 대응 서비스 개발

- 원산지 조사의 효율성 제고 방안 제시. 올바른 원산지 선순환 문화 정착
- 단속원들이 현장에서 업무용 스마트폰을 통해 활용할 수 있는 단속 가이드 시스템 개발
- 본인만의 노하우를 반영한 단속추천 지역을 분석. 지도상으로 시각화 할 수 있는 단속 가이드 시스템 개발

<그림 1-4> 연구 진행 방향



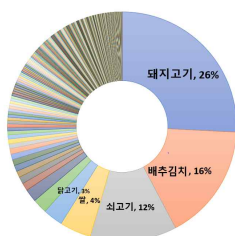
12

### 3 연구 범위와 연구 방법

#### 가. 연구 범위

- 본 연구는 위반이 발생하는 지역적 요인 등의 도출에 관련한 문헌연구와 시중에 유통되고 있는 수입 물품의 원산지 표시 위반 조사기관에 대한 현장방문과 내부 공공데이터를 포함한 빅데이터 분석을 병행하였음
- 분석지역은 조사가 진행된 6개 도시(서울특별시, 부산광역시, 대구광역시, 인천광역시, 대전광역시, 광주광역시)로 선정함
- 전체 단속품목의 61%로 큰 비중을 차지하고 있는 5개 품목으로 조사 범위를 한정
- 분석 기간은 2009년부터 2018년까지 10년 동안의 데이터를 가지고 진행을 함
- 분석품목은 돼지고기, 배추김치, 쇠고기, 쌀, 닭고기 등 5가지 품목을 주로 진행함<sup>1)</sup>

<그림 1-5> 전체 품목 대비 상위 대비 상위품목 비율



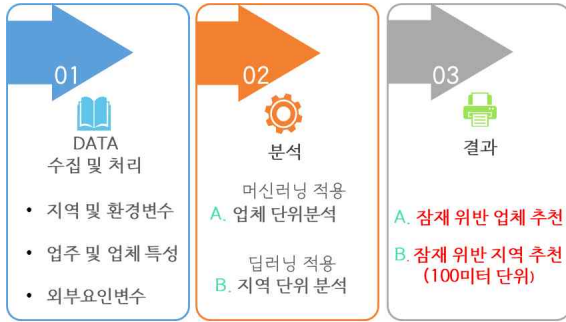
1) 국립농산물품질관리원, 2015 농산물품질관리연보, 농림수산물부(2016), 117쪽

13

## 나. 연구 방법

- 시기·지역·품목 세 가지를 모두 고려한 분석을 위해 지역·환경, 업주·업체특성, 외부요인 변수 등 업체추천의 경우 24개, 지역추천의 경우 55개 변수의 데이터를 수집·가공함
- 업체추천을 위해서는 머신러닝 알고리즘 8개로 분석을 진행하였으며 그중 가장 높은 정확도를 가지는 모형으로 추천업체 분석을 진행함
- 지역 단위추천을 위해서는 딥러닝 알고리즘 중 CNN 알고리즘을 사용하여 100m<sup>2</sup> 그리드 단위로 분석을 진행함
- 최종적으로 잠재 위반가능성이 높은 업체들을 추천하고 잠재 위반가능성이 높은 지역을 추천하는 시스템을 개발함

<그림 1-6> 연구개요



## II 인공지능 추천서비스 사례연구

### 1 국내사례

#### □ 검역 정보 분석 알고리즘<sup>1)</sup>, 한국

- 검역 정보 분석 알고리즘은 출입국가 증가와 가축 전염병 발생 등 국민의 건강과 안전을 위협하는 요인을 탐색함
  - 인천국제공항의 여행객이나 보따리상에 의한 밀반입 적발건수는 2015년 이후 매년 12% 이상 증가하고 있고 적발에 따른 과태료부과 역시 매년 30% 증가하는 추세로 사전예방 중심의 과학적 국경 검역 기반 구축
- 해당 알고리즘은 검역 본부로부터 인천국제공항의 과거 3년간(2015 ~ 2017년)여객기 입항정보(약 50만건), 검역·단속 현황(약 37만건), 탐지견 운영현황 등을 제공받아 기계학습으로 분석하여 검역 대항 항공기를 추천하고 전체 검역 실적의 30%를 차지하는 탐지견 운영의 효율적 방안을 제시함
- 인공지능 알고리즘을 활용하여 휴대 물품 밀반입 가능성이 높은 항공기를 선별하고 불법 휴대 물품 반입 위험도를 예측함.
  - 5종의 인공지능 알고리즘을 테스트하여 최적의 예측모형(Random Forest 분류)을 선정하고 데이터 학습 및 검증 과정을 수차례 반복하여 높은 정확도(92.4%)의 인공지능 예측모형을 구축함

1) 김세정, 출입국가 8천만명 시대...인공지능으로 밀반입 차단 2018년 11월 6일 등록, 2019년 11월 20일 접속, <http://livesnews.com/news/article.html?no=27088>

## 가. PredPol, 미국

- 프레드폴(PredPol)은 2008년 미국 LA 경찰청(LAPD)에서 특정 범죄의 예방을 위해 해당 범죄에 대한 예측가능성을 제시함으로써 예측적 경찰 활동(Predictive Policing)을 제안하는 시스템임
- 캘리포니아의 산타크루즈 지역의 인구는 증가한 반면 경찰력은 감소하였으며 인력보강에 대한 계획이 없었음
- 또한 범죄는 증가하는 상황으로 이에 대한 대책으로 효과적이고 효율적인 범죄대응 전략이 필요해짐. 프레드폴의 개발과 함께 미국 최초로 예측적 경찰활동 시행함
- 프레드폴은 기계학습 알고리즘을 이용하여 범죄 유형, 범죄 발생 위치와 시간에 대해 예측하고 순찰 경로추천 등의 기능을 제공함
- 범죄 유형과 발생 장소, 발생 시간 3가지 측면에서 150\*150미터의 격자지도에 잠재적 범죄 발생률을 표시해 줌으로써 특정 시간대에 범죄가 증가할 것으로 예상되는 지점을 지속적으로 추천함
- 실시간 업데이트를 통해 근무교대 전이나 근무 중에 정보를 제공하고 기존 경찰관들의 경험과 직관을 보완하는 역할로 신입 경찰관들의 직무수행 능력을 향상시키는 기여를 함
- 프레드폴은 범행이 치밀하게 계획되어지는 침입 절도의 예측에 있어 특히 효과적임이 확인되었으며 계속되는 시스템의 조율과정을 통해 효과성이 지속되는 것으로 드러남

나. 파일럿(PILOT)<sup>1)</sup>, 미국

- PILOT은 “정보기반 개입대상 예측”을 목적으로 범죄가 주로 발생하는 핫스팟(Hot Spot)을 예측하는 시스템임
- 먼저 범죄와 무질서에 관련된 최근 자료를 바탕으로 범죄증가가 눈에 띄는 지역을 1개월간 관찰하여 강도·침입절도·차량털이·거리절도·차량 절도를 “전술적 개입대상 범죄(tactical crime)”로 규정하고, 이들을 예측하기 위해 911 신고 전화·경범죄·청소년 검거기록·계절적 지표 등을 활용함. 이와 같은 변수들을 활용하여 개입대상 범죄들에 대해서는 시차변수와 공간시차 변수를 도출함
- 약 37제곱 미터(400제곱피트) 격자 단위로 구축된 데이터를 이용해 두 가지 예측 기법을 적용
- 먼저 로지스틱 회귀분석 모델을 사용하여 범죄가 발생할 가능성을 각 지역과 시간 단위로 분석하여 시간적·공간적으로 근접한 수치들의 군집을 도출함. 이를 통해 특정 지역에서 발생하는 특정 현상이 범죄가 발생한 인접한 시간대나 이웃한 지역과 유사한지를 판단함
- 두 번째 기법은 위험지역 모델링(RTM)으로 범죄와 관련이 큰 지리적·공간적 특성(예: 상대적으로 많은 수감자나 가석방자 수, 과거 범죄나 무질서 다발지역 등)을 격자 단위로 분석함. 범죄 발생에 영향을 주는 특성이 하나 발견될 때마다 1점을 매기는 방식으로 각 셀의 점수를 합산함으로써 높은 점수가 부여된 셀이 고위험 지역으로 분류
- 로지스틱 회귀분석과 위험지역 모델링, 두 가지 분석결과를 지도에 동시에 표시해서 제공함으로써 범죄 예방성 방법활동을 지원함

1) 임운식, 스마트 치안과 관련된 해외사례, 2018년 8월 6일 등록,  
2019년 11월 20일 접속, <https://knpo.police.ac.kr/pds/1539934763318.pdf>

다. DAS(Domain Awareness System)<sup>1)</sup>, 미국

- NYPD의 Domain Awareness System(DAS)은 대표적인 스마트 경찰활동 사례로 공공안전 및 보안을 유지하고 테러 활동을 탐지하고 예방하기 위해 개발된 시스템으로 NYPD는 뉴욕시 곳곳에 위치하고 있음
- 예측적 경찰활동이 미래를 예측하는 것에 초점을 맞추는데 반해 DAS는 도시 내 데이터 소스를 통해 발생하는 엄청난 양의 데이터를 처리해 “현재”를 명확하게 보여줌
- NYPD는 2백대 이상의 자동차 번호판 탐지기와 2천대의 방사선 센서 그리고 경찰 데이터베이스를 활용하여 도시 내 운행 중인 모든 차량의 정보를 실시간으로 수집·분석·감시하며 실시간으로 지도상에 매핑·제공함으로써 범죄대응에 일조함
- DAS는 이전의 개별적 분석과는 다르게 사람·사물·장소 간 연관성을 인공지능으로 분석해 버려진 가방과 같은 위협까지도 감지함
  - 용의자와 연관된 차량이 현재 어디에 있는지 과거 수개월 동안 어디에 있었는지를 실시간으로 조회·추적할 수 있음

1) 한국형사정책연구원, 범죄 빅데이터를 활용한 범죄예방시스템 구축을 위한 예비연구, 2015년 12월 등록, 2019년 11월 20일 접속,  
[http://img.kisti.re.kr/tr\\_img/2016002/rtrtko00000232862.pdf](http://img.kisti.re.kr/tr_img/2016002/rtrtko00000232862.pdf)

라. HART<sup>1)</sup>, 영국

- HART(Harm Assessment Risk Tool)는 영국의 더럼에서 사용되는 AI 기반 치안 플랫폼으로 재판 전 석방과 가석방 데이터를 활용해 범죄자가 사회에 해가 될 가능성을 평가하는 프로그램임. 범죄자의 재범 가능성을 낮음(Low), 보통(Medium), 높음(High)으로 분류하고 각각 다른 관련 정보를 제공함
- HART는 연령, 성별 및 범죄 기록을 포함하여 34가지 범주의 데이터를 사용하여 사람들을 위험이 낮거나 중간 정도 또는 높은 위험으로 평가함. 우편번호를 기준으로 지속적으로 데이터를 업데이트함은 물론 알고리즘 역시 조율됨
- 이에 HART는 영국 경찰이 가장 먼저 사용하는 알고리즘 중 하나로 사용됨. 피의자를 수감해야 하는지 여부를 결정하지는 않지만 사람이 Checkpoint라는 재할 프로그램에 의뢰되어야 하는지 또 추가적인 감시가 필요한지 등 경찰관의 선택을 보조함

1) 범준성, 박원주, 신지호, 이용태, 지능형 치안 서비스 기술 동향, 2019년 2월 등록, 2019년 11월 20일 접속,  
[https://ettrends.etri.re.kr/ettrends/175/0905175010/34-1\\_111-122.pdf](https://ettrends.etri.re.kr/ettrends/175/0905175010/34-1_111-122.pdf)



<표 2-1> 해외사례

사례명 (사용국가, 도입년월)	내 용	예측 모형
프레드폴 (PredPol)  미국 (2008.11)	<ul style="list-style-type: none"> <li>- 유형, 위치, 시간 데이터만을 사용하여 최근 범죄일수록 가중치를 주고 시차 변수, 공간 시차 변수, <math>37m^2</math> 격자 구분</li> <li>· 기법1: 격자별 범죄확률 계산(로지스틱 회귀분석)</li> <li>· 기법2: 범죄와 관련이 큰 지리적·공간적 특성을 하나당 1점씩 부여하여 점수 합산 후 <math>150*150</math> 격자지도에 미래 발생확률을 표시함으로 침입절도(재산범죄)에 효과(=합리적 계획에 효과)</li> <li>- 예측지도는 매달 업데이트 제공</li> </ul>	<ul style="list-style-type: none"> <li>· 격자단위 분석</li> <li>· 지리·공간정보 분석</li> <li>· 로지스틱 회귀 분석</li> <li>· 히트맵</li> </ul>
하트 (HART)  영국 (2017.03)	<ul style="list-style-type: none"> <li>- HART는 R 프로그래밍 언어를 사용하고 일련의 서로 다른 결과를 기반으로 예측하는 방법인 랜덤 포레스트를 통해 의사결정을 하는 머신러닝 시스템</li> <li>- 이 모형은 범죄자 '위험성'을 예측하고 고위험군, 중급 또는 저위험군으로 분류</li> <li>- HART가 내리는 모든 결정은 과거 데이터를 기반으로 하며, 이전 정보를 보고 미래의 결과를 예측함 (예: 첫 번째 모형에서 2008년부터 2012년까지 104,000건의 양육권 사건에 대한 세부사항을 시스템에 제공했고, 여기에서 위치데이터를 포함하여 34명의 예측자료를 사용하여 각 사람에 대해 예측함)</li> <li>- HART가 도달한 모든 결론은 시스템의 509표를 기반으로 함(투표는 낮음, 보통 또는 높음)</li> </ul>	<ul style="list-style-type: none"> <li>· 랜덤 포레스트</li> <li>· 로지스틱 회귀 분석</li> </ul>

20

파일럿 (PILOT)  미국 (2013.11)	<ul style="list-style-type: none"> <li>- 루이지애나 Shreveport 경찰서의 PILOT은 범죄 외의 정보도 사용하고 있는 핫스팟 예측 및 예방 시스템</li> <li>- 위험지역 모델링(RTM)으로서 먼저 각 셀에 분포하는 범죄와 관련이 큰 지리적·공간적 특성</li> <li>- 범죄가 발생할 가능성(확률)을 각 셀(격자로 구분된 개별 구역), 시차(lag, 시간적·공간적으로 근접한 수치들의 평균) 두 가지 방식을 사용</li> </ul>	<ul style="list-style-type: none"> <li>· 격자단위 분석</li> <li>· 지리·공간정보 분석</li> <li>· 로지스틱 회귀 분석</li> <li>· 히트맵</li> </ul>
Domain Awareness System(DAS)  미국 (2012.06)	<ul style="list-style-type: none"> <li>- DAS는 사람·사물·장소 간 연관성을 밝혀, 비디오 분석 소프트웨어(CNN)을 이용해서 위협을 감지</li> <li>- 자동차 번호판 탐지기를 통해 1,600만 건의 번호판 데이터베이스를 가지고, 여기에는 위치정보가 포함</li> <li>- 2백대 이상의 자동차 번호판 탐지기·2천대의 방사선 센서 그리고 경찰 데이터베이스의 정보를 실시간으로 수집하고 분석</li> <li>- 또한 매핑 기능이 추가되어 있어 경찰이 과거보다 훨씬 쉽게 상황을 이해하는데 도움을 줌</li> </ul>	<ul style="list-style-type: none"> <li>· 로지스틱 회귀 분석</li> <li>· CNN</li> </ul>

21

### Ⅲ 원산지 조사 및 위반 단속 분석

#### 1 원산지 조사 및 단속 분석

##### 가. 전국 현황

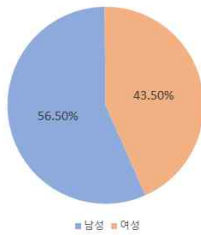
##### 1) 단속 업주별 성별

<표 3-1> 단속 업주별 성별

기간: 2009 ~ 2018년

성별	단속 업주(명)	비율(%)
여성	7,420	43.5
남성	9,621	56.5

<그림 3-1> 단속 업주별 성별



##### 2) 단속 업주 나이

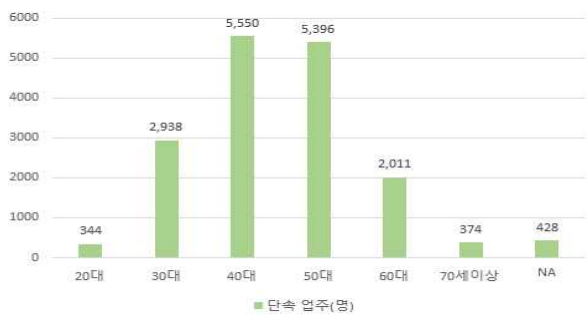
<표 3-2> 단속 업주 나이

기간: 2009 ~ 2018

나이범주	단속 업주(명)	비율(%)
20대	344	2.0
30대	2,938	17.2
40대	5,550	32.6
50대	5,396	31.7
60대	2,011	11.8
70대 이상	374	2.2
NA	428	2.5

<그림 3-2> 단속 업주 나이

(기간: 2009 ~ 2018)



### 3) 지역별 단속 업주 나이

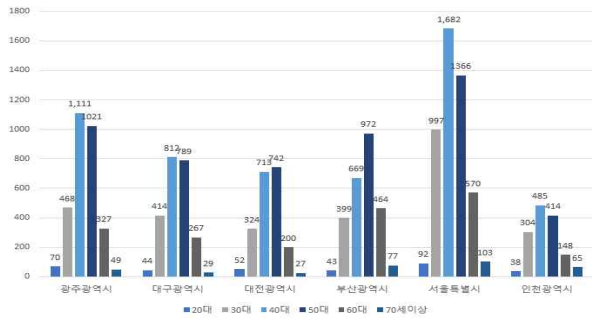
<표 3-3> 지역별 단속 업주 나이

기간: 2009 ~ 2018

적발사유	20대		30대		40대		50대		60대		70대 이상	
	(명)	(%)	(명)	(%)	(명)	(%)	(명)	(%)	(명)	(%)	(명)	(%)
광주광역시	70	2.3	468	15.4	1,111	36.5	1,021	33.5	327	10.7	49	1.6
대구광역시	44	1.9	414	17.6	812	34.5	789	33.5	267	11.3	29	1.2
대전광역시	52	2.5	324	15.7	713	34.7	742	36.1	200	9.7	27	1.3
부산광역시	43	1.6	399	15.2	669	25.5	972	37.0	464	17.7	77	2.9
서울특별시	92	1.9	997	20.7	1,682	35.0	1,366	28.4	570	11.9	103	2.1
인천광역시	38	2.6	304	20.9	485	33.4	414	28.5	148	10.2	65	4.5

<그림 3-3> 지역별 단속 업주 나이

(기간: 2009 ~ 2018)



#### 나. 대도시별 단속 현황

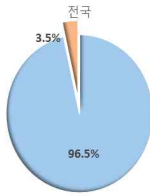
- 2009년부터 10년간 원산지 표시 조사대상 48만여 개소를 조사하여 위반 업소 16,769건을 단속하였으며 전국적으로 평균 약 3.5%의 단속률을 보임
- 3장의 원산지 조사 및 위반 단속 분석에 사용한 데이터는 농관원에서 제공한 전국업체 단위의 데이터를 그대로 사용함. 그러나 머신러닝 및 딥러닝의 모형에 사용된 데이터도 농관원에서 제공한 기존 전국 업체 단위 데이터를 사용하였으나, 데이터의 주소 정제 과정에서 누락된 데이터가 있어 일부를 제외하고 사용하였음. 따라서 기존 전국 업체 단위 데이터 단속률과 머신러닝과 딥러닝의 모형을 위해 정제한 데이터를 사용한 단속률 간에 차이가 발생됨
- 대도시별 단속률은 대구광역시가 6.2%로 가장 높게 나타났으며, 광주광역시(6.1%), 대전광역시(3.7%), 부산광역시(3.1%), 서울특별시(2.9%), 인천광역시(2.0%) 순으로 높음

<표 3-4> 대도시별 단속 현황

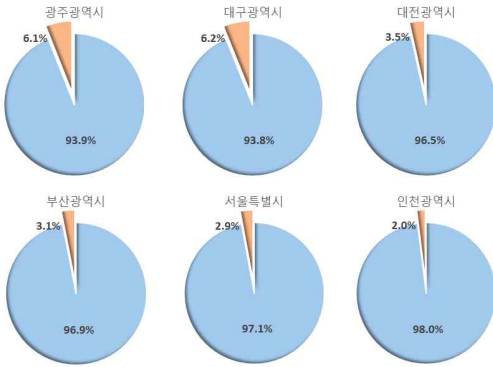
단위: 건, 기간: 2009 ~ 2018

시도	전체	단속	미단속	단속률(%)
전국	480,165	16,769	463,396	3.5
광주광역시	51,064	3,091	47,973	6.1
대구광역시	39,783	2,451	37,332	6.2
대전광역시	57,560	2,109	55,451	3.7
부산광역시	87,327	2,686	84,641	3.1
서울특별시	166,545	4,890	161,655	2.9
인천광역시	77,886	1,542	76,344	2.0

<그림 3-4> 단속률



<그림 3-5> 주요 도시 단속률  
(기간: 2009 ~ 2018)



#### 다. 연도별 단속 비율

○ 2018년 전국 단속 비율은 2.5%로서 전년도 대비 1.4%p 감소함

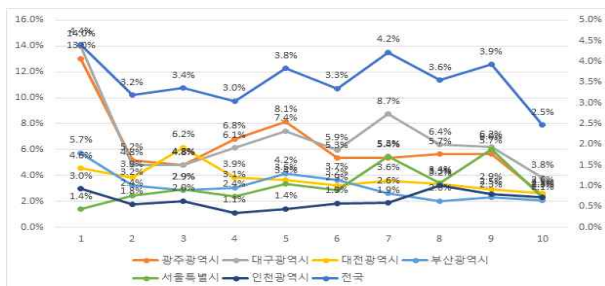
<표 3-5> 연도별 단속 비율

단위: %, 기간: 2009 ~ 2018

시도	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	평균
전국	4.4	3.2	3.4	3.0	3.8	3.3	4.2	3.6	3.9	2.5	3.5
광주광역시	13.0	5.2	4.8	6.8	8.1	5.3	5.4	5.7	5.7	2.5	6.1
대구광역시	14.0	4.8	4.8	6.1	7.4	5.9	8.7	6.4	6.2	3.8	6.2
대전광역시	4.6	3.9	6.2	3.9	3.6	3.2	3.6	3.3	2.9	2.6	3.7
부산광역시	5.7	3.2	2.9	3.1	4.2	3.7	2.6	2.0	2.3	2.1	3.1
서울특별시	1.4	2.4	2.9	2.4	3.4	2.9	5.5	3.4	6.0	2.3	3.0
인천광역시	3.0	1.8	2.0	1.1	1.4	1.9	1.9	3.2	2.5	2.3	2.0

<그림 3-6> 연도별 단속 비율

(기간: 2009 ~ 2018)



라. 연도별 단속 건수

- 2018년 한해 동안 전국적으로 원산지 표시대상 53,330개소를 조사하여 위반 업소 1,318개소를 적발하였음

<표 3-6> 연도별 단속 건수

단위: 건, 기간: 2009 ~ 2018

시도	구분	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
전국	조사 건수	51,735	58,223	55,882	51,606	41,339	47,230	39,701	43,032	38,087	53,330
	단속 건수	2,281	1,856	1,876	1,570	1,587	1,580	1,677	1,528	1,496	1,318
광주광역시	조사 건수	5,478	6,871	6,574	3,817	3,602	5,335	5,573	3,855	3,101	6,858
	단속 건수	713	357	317	261	293	285	300	218	176	171
대구광역시	조사 건수	1,919	5,835	5,845	4,397	3,679	3,840	2,725	3,702	3,226	4,615
	단속 건수	268	280	280	270	274	228	238	236	200	177
대전광역시	조사 건수	6,958	6,445	3,555	5,193	4,681	5,449	5,408	5,527	7,014	7,330
	단속 건수	317	249	219	200	170	176	195	184	205	194
부산광역시	조사 건수	8,270	9,893	9,631	8,154	6,214	6,670	8,013	9,209	10,409	10,864
	단속 건수	474	320	276	250	258	244	210	187	240	227
서울특별시	조사 건수	22,341	20,098	19,488	20,270	13,570	16,591	11,119	16,319	8,970	17,779
	단속 건수	307	488	566	479	456	474	606	561	539	414
인천광역시	조사 건수	6,769	9,081	10,789	9,775	9,593	9,345	6,863	4,420	5,367	5,884
	단속 건수	202	162	218	110	136	173	128	142	136	135

1) 전국 연도별 단속 내역

- 단속 내역은 거짓표시와 미표시가 지속적으로 대부분의 비중을 차지하고 있으며 지역별로 2009년의 광주광역시와 부산광역시를 제외하고 1) 비슷한 비율로 관측됨
- 2018년 단속된 업체 1,318개소 중 거짓표시는 829건(62.9%)으로 거짓 표시 비율이 가장 높았으며 미표시는 417건(31.6%)으로 나타남

<표 3-7> 전국 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
거짓표시	1,032	1,170	1,169	852	995	965	955	924	883	829
미표시	1,184	578	548	560	486	494	583	476	503	417
손상변경	-	-	-	1	1	-	-	1	-	-
시정명령위반	-	-	-	2	1	-	2	5	3	-
영수증미비치	-	-	7	-	-	-	2	2	3	1
위반행위 방치	-	-	-	1	1	-	-	-	-	-
위장판매	12	20	26	15	27	28	45	48	37	24
조사거부	-	-	-	-	-	-	1	-	-	-
표시방법위반	4	4	-	23	9	6	19	2	20	5
거짓표시	-	1	-	-	-	-	-	-	-	-
혼동우려	49	82	120	114	66	84	69	66	42	40
혼합판매	-	1	6	2	1	3	1	4	5	2

<그림 3-7> 전국 연도별 단속 내역

(기간: 2009 ~ 2018)



- 1) 미표시의 단속 건수가 거짓표시보다 많음

## 2) 광주광역시 연도별 단속 내역

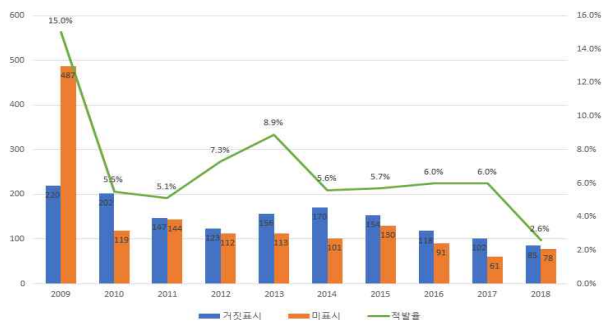
<표 3-8> 광주광역시 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
거짓표시	220	202	147	123	156	170	154	118	102	85
미표시	487	119	144	112	113	101	130	91	61	78
영수증미비치	-	-	-	-	-	-	1	-	3	1
위장판매	2	3	3	1	2	2	-	1	-	-
표시방법위반	-	-	-	3	1	-	12	-	4	-
혼동우려	4	33	23	22	21	12	3	8	6	7

<그림 3-8> 광주광역시 연도별 단속 내역

(기간: 2009 ~ 2018)



## 3) 대구광역시 연도별 단속 내역

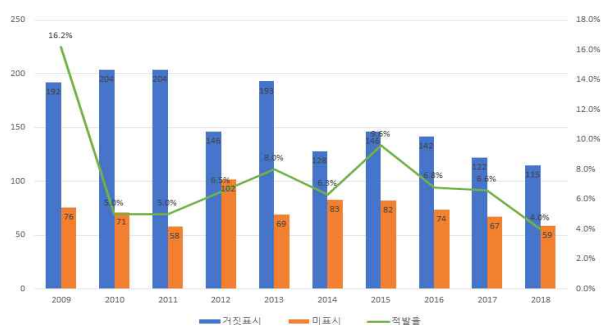
<표 3-9> 대구광역시 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
거짓표시	192	204	204	146	193	128	146	142	122	115
미표시	76	71	58	102	69	83	82	74	67	59
손상변경	-	-	-	1	-	-	-	-	-	-
시정명령위반	-	-	-	-	-	-	-	2	-	-
영수증미비치	-	-	2	-	-	-	-	-	-	-
위반행위 방치	-	-	-	1	1	-	-	-	-	-
위장판매	-	-	2	3	1	6	1	1	1	2
표시방법위반	-	2	-	-	-	-	2	-	1	-
혼동우려	-	3	13	17	10	11	7	17	9	1
혼합판매	-	-	1	-	-	-	-	-	-	-

<그림 3-9> 대구광역시 연도별 단속 내역

(기간: 2009 ~ 2018)



#### 4) 대전광역시 연도별 단속 내역

<표 3-10> 대전광역시 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
거짓표시	148	166	144	116	94	98	113	99	124	131
미표시	125	72	56	53	59	52	74	75	70	60
혼동우려	43	7	16	30	14	25	4	9	7	3
표시방법위반	1	1	-	-	2	-	-	-	4	-
위장판매	-	2	3	1	1	1	4	1	-	-
거짓표시	-	1	-	-	-	-	-	-	-	-

<그림 3-10> 대전광역시 연도별 단속 내역

(기간: 2009 ~ 2018)



#### 5) 부산광역시 연도별 단속 내역

<표 3-11> 부산광역시 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
미표시	322	118	94	116	90	87	71	50	82	76
거짓표시	148	179	169	125	154	151	130	126	144	147
위장판매	3	6	2	2	-	-	-	3	-	1
혼동우려	1	16	9	7	11	2	7	3	2	1
손상변경	-	-	-	-	1	-	-	-	-	-
영수증미비치	-	-	2	-	-	-	1	2	-	-
표시방법위반	-	-	-	-	2	2	-	-	7	-
혼합판매	-	1	-	-	-	2	1	3	5	2

<그림 3-11> 부산광역시 연도별 단속 내역

(기간: 2009 ~ 2018)



## 6) 서울특별시 연도별 단속 내역

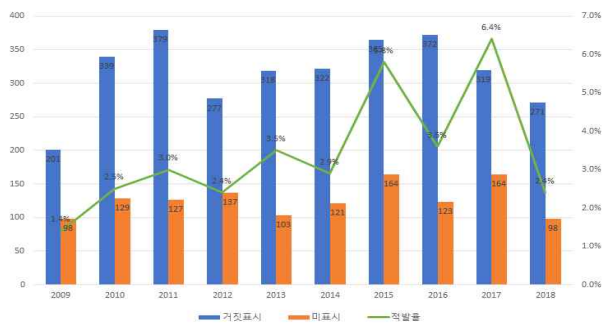
<표 3-12> 서울특별시 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
거짓표시	201	339	379	277	318	322	365	372	319	271
미표시	98	129	127	137	103	121	164	123	164	98
손상변경	-	-	-	-	-	-	-	1	-	-
시정명령위반	-	-	-	2	1	-	2	3	3	-
영수증미비치	-	-	3	-	-	-	-	-	-	-
위장판매	5	9	10	8	23	17	37	36	32	12
조사가부	-	-	-	-	-	-	1	-	-	-
표시방법위반	3	1	-	20	4	3	5	2	4	5
혼동우려	-	10	46	34	6	11	32	24	17	28
혼합판매	-	-	1	1	1	-	-	-	-	-

<그림 3-12> 서울특별시 연도별 단속 내역

(기간: 2009 ~ 2018)



## 7) 인천광역시 연도별 단속 내역

<표 3-13> 인천광역시 연도별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
거짓표시	123	80	126	65	80	96	47	67	72	80
미표시	76	69	69	40	52	50	62	63	59	46
위장판매	2	-	6	-	-	2	3	6	4	9
혼동우려	1	13	13	4	4	23	16	5	1	-
표시방법위반	-	-	-	-	-	1	-	-	-	-
혼합판매	-	-	4	1	-	1	-	1	-	-

<그림 3-13> 인천광역시 연도별 단속 내역

(기간: 2009 ~ 2018)





## 마. 월별 단속 비율

### 1) 월별 단속 비율

- 월별 단속 비율은 1, 2월에 높고 이후 감소하다 9월에 다시 증가하는 추세를 보이며 대구광역시의 1월 단속 비율은 47.6%로 매우 높게 나타남

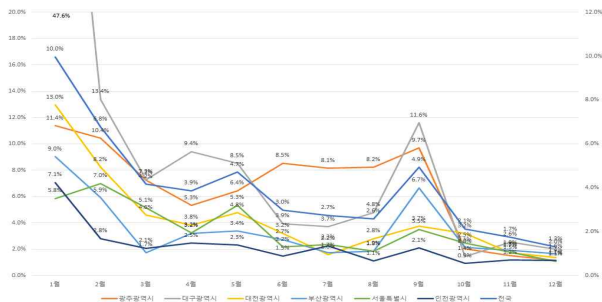
<표 3-14> 월별 단속 비율

단위: %, 기간: 2009 ~ 2018

시도	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	10.0	6.8	4.2	3.9	4.7	3.0	2.7	2.6	4.9	2.1	1.7	1.3
광주광역시	11.4	10.4	7.2	5.3	6.4	8.5	8.1	8.2	9.7	2.0	1.5	1.1
대구광역시	47.6	13.4	7.3	9.4	8.5	3.9	3.7	4.8	11.6	1.4	2.6	2.0
대전광역시	13.0	8.2	4.6	3.8	4.8	3.2	1.6	2.8	3.7	3.2	1.7	1.4
부산광역시	9.0	5.9	1.7	3.2	3.4	2.7	1.7	1.8	6.7	2.1	1.9	1.6
서울특별시	5.8	7.0	5.1	3.2	5.3	2.2	2.3	1.9	3.5	2.5	1.8	1.1
인천광역시	7.1	2.8	2.1	2.5	2.3	1.5	2.2	1.1	2.1	0.9	1.2	1.1

<그림 3-14> 월별 단속 비율

(기간: 2009 ~ 2018)



### 2) 전국 월별 단속 내역

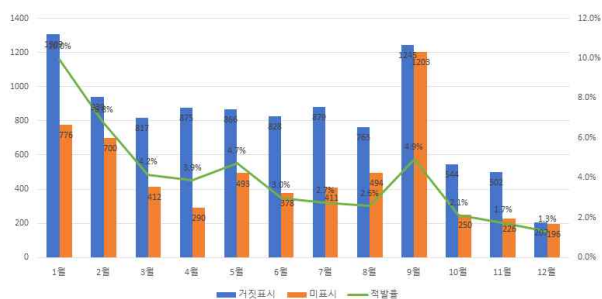
<표 3-15> 전국 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	1,309	939	817	875	866	828	879	763	1,245	544	502	207
미표시	776	700	412	290	493	378	411	494	1,203	250	226	196
손상변경	-	1	-	1	-	-	-	-	1	-	-	-
시정명령위반	1	6	1	1	-	-	-	-	1	2	-	1
영수증미비치	-	4	3	2	1	2	-	-	2	-	-	1
위반행위 방지	1	-	-	-	-	-	-	-	-	-	-	1
위장판매	39	40	11	44	21	21	30	25	30	10	6	5
조사거부	-	-	-	-	-	-	-	1	-	-	-	-
표시방법위반	2	1	6	5	12	8	5	11	24	5	13	-
거짓표시	-	-	-	-	-	-	-	1	-	-	-	-
혼동우려	87	64	71	50	71	55	71	89	63	56	38	17
혼합판매	3	1	3	1	3	3	3	2	-	3	2	1

<그림 3-15> 전국 월별 단속 내역

(기간: 2009 ~ 2018)



### 3) 광주광역시 월별 단속 내역

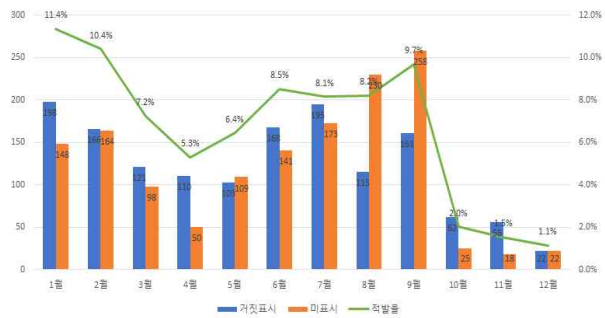
<표 3-16> 광주광역시 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	198	166	121	110	103	168	195	115	161	62	56	22
미표시	148	164	98	50	109	141	173	230	258	25	18	22
영수증미비치	-	1	-	-	1	2	-	-	1	-	-	-
위장판매	4	1	-	2	1	-	2	2	1	-	1	-
표시방법위반	-	-	-	-	5	1	-	1	10	-	3	-
혼동우려	19	8	16	11	13	12	18	17	18	2	5	-

<그림 3-16> 광주광역시 월별 단속 내역

(기간: 2009 ~ 2018)



### 4) 대구광역시 월별 단속 내역

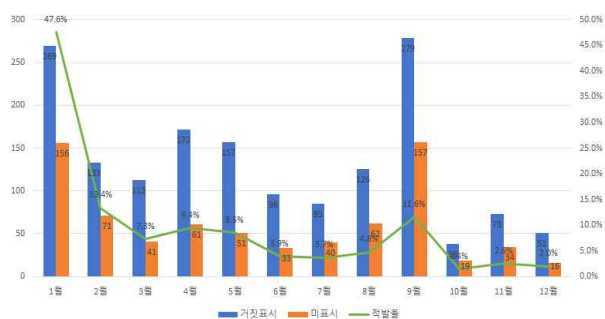
<표 3-17> 대구광역시 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	269	133	113	172	157	96	85	126	279	38	73	51
미표시	156	71	41	61	51	33	40	62	157	19	34	16
손상변경	-	-	-	1	-	-	-	-	-	-	-	-
시정명령위반	-	2	-	-	-	-	-	-	-	-	-	-
영수증미비치	-	-	1	1	-	-	-	-	-	-	-	-
위반행위 방치	1	-	-	-	-	-	-	-	-	-	-	1
위장판매	2	2	3	3	1	-	1	-	2	2	1	-
표시방법위반	-	-	-	-	-	-	1	1	2	-	1	-
혼동우려	17	9	2	5	9	5	7	11	12	1	6	4
혼합판매	-	-	-	-	-	-	-	-	-	-	1	-

<그림 3-17> 대구광역시 월별 단속 내역

(기간: 2009 ~ 2018)



## 5) 대전광역시 월별 단속 내역

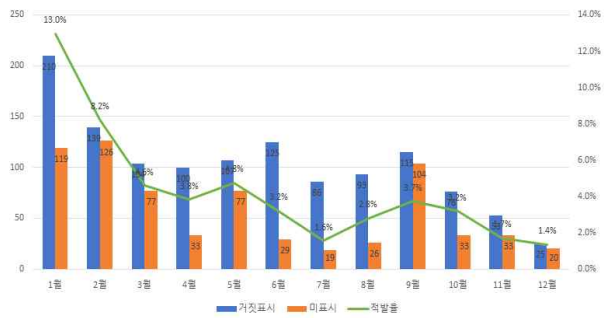
<표 3-18> 대전광역시 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	210	139	104	100	107	125	86	93	115	76	53	25
미표시	119	126	77	33	77	29	19	26	104	33	33	20
위장판매	1	6	1	-	-	-	-	1	1	1	-	2
표시방법위반	-	-	-	-	-	4	-	-	1	1	2	-
거짓표시	-	-	-	-	-	-	-	1	-	-	-	-
혼동우려	17	12	6	11	18	14	6	41	11	21	1	-

<그림 3-18> 대전광역시 월별 단속 내역

(기간: 2009 ~ 2018)



## 6) 부산광역시 월별 단속 내역

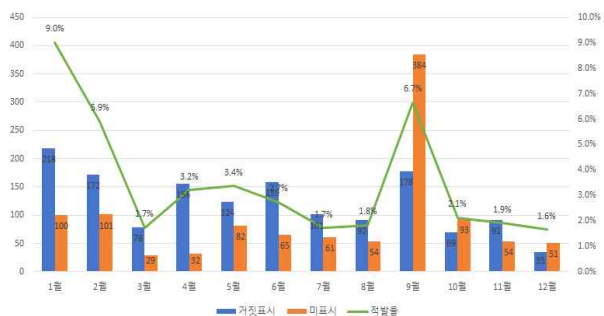
<표 3-19> 부산광역시 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	218	172	78	156	124	159	101	92	178	69	91	35
미표시	100	101	29	32	82	65	61	54	384	93	54	51
손상변경	-	1	-	-	-	-	-	-	-	-	-	-
영수증미비치	-	2	-	1	-	-	-	-	1	-	-	1
위장판매	3	3	-	3	1	2	-	-	1	3	1	-
표시방법위반	-	-	-	-	6	-	-	-	-	1	4	-
혼동우려	10	5	5	2	3	6	6	8	3	5	5	1
혼합판매	1	1	2	1	1	2	3	-	-	2	1	-

<그림 3-19> 부산광역시 월별 단속 내역

(기간: 2009 ~ 2018)



## 7) 서울특별시 월별 단속 내역

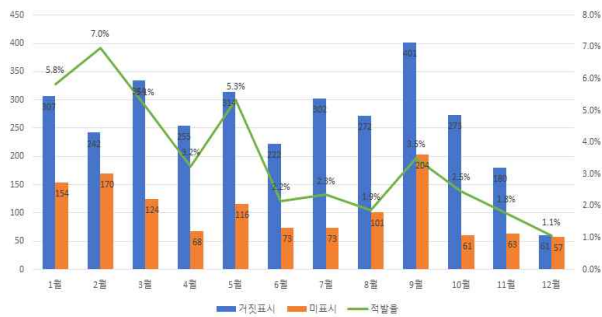
<표 3-20> 서울특별시 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	307	242	334	255	314	222	302	272	401	273	180	61
미표시	154	170	124	68	116	73	73	101	204	61	63	57
손상변경	-	-	-	-	-	-	-	-	1	-	-	-
시정명령위반	1	4	1	1	-	-	-	-	1	2	-	1
영수증미비치	-	1	2	-	-	-	-	-	-	-	-	-
위장판매	22	23	7	34	17	17	21	16	24	4	1	3
조사가부	-	-	-	-	-	-	-	1	-	-	-	-
표시방법위반	2	1	6	5	1	3	3	9	11	3	3	-
혼동우려	13	14	31	16	25	10	25	8	16	23	16	11
혼합판매	-	-	1	-	-	-	-	-	-	1	-	1

<그림 3-20> 서울특별시 월별 단속 내역

(기간: 2009 ~ 2018)



## 8) 인천광역시 월별 단속 내역

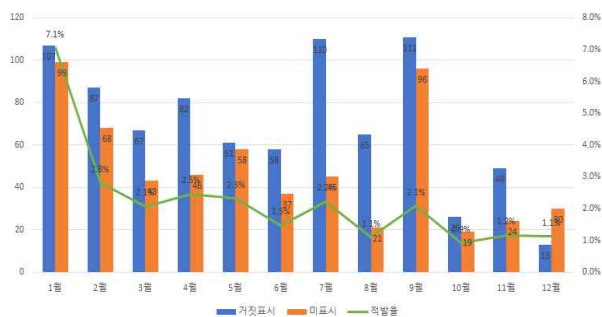
<표 3-21> 인천광역시 월별 단속 내역

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
거짓표시	107	87	67	82	61	58	110	65	111	26	49	13
미표시	99	68	43	46	58	37	45	21	96	19	24	30
위장판매	7	5	-	2	1	2	6	6	1	-	2	-
표시방법위반	-	-	-	-	-	-	1	-	-	-	-	-
혼동우려	11	16	11	5	3	8	9	4	3	4	5	1
혼합판매	2	-	-	-	2	1	-	2	-	-	-	-

<그림 3-21> 인천광역시 월별 단속 내역

(기간: 2009 ~ 2018)



바. 요일별 단속 현황

1) 전국 요일별 단속 비율

- 6대 도시의 요일별 단속률은 전반적으로 토요일에 높았으며 일요일에는 낮은 단속률을 보이고 있으나 대전은 일요일에 단속률이 높았음
- 6대 도시별 요일별 단속 내역의 주요 현황은 아래와 같음

<표 3-22> 전국 요일별 단속 현황

단위: %, 기간: 2009 ~ 2018							
요일	월	화	수	목	금	토	일
전국	3.5	3.5	3.5	3.5	3.0	9.8	4.3
광주광역시	5.1	5.1	6.6	8.0	5.9	6.5	3.4
대구광역시	6.4	7.2	6.2	5.6	4.9	18.0	12.8
대전광역시	3.3	3.4	4.1	3.7	3.4	23.3	30.2
부산광역시	2.8	3.9	3.5	2.9	2.2	4.5	1.6
서울특별시	3.4	2.9	2.4	2.7	2.7	9.8	2.6
인천광역시	1.8	1.7	2.5	1.9	1.5	26.8	16.5

<그림 3-22> 전국 요일별 단속 현황

(기간: 2009 ~ 2018)



2) 서울특별시 요일별 단속 현황

- 서울은 토요일에 9.8%로 가장 높았고, 수요일에 2.4%로 가장 낮았음

<그림 3-23> 서울특별시 요일별 단속 현황

(기간: 2009 ~ 2018)

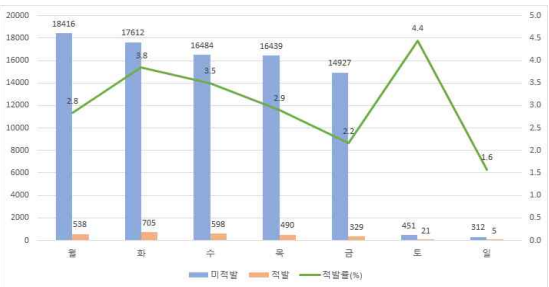


3) 부산광역시 요일별 단속 현황

- 부산은 토요일에 4.4%로 가장 높았고, 일요일에 1.6%로 가장 낮았음

<그림 3-24> 부산광역시 요일별 단속 현황

(기간: 2009 ~ 2018)



#### 4) 인천광역시 요일별 단속 현황

- 인천광역시는 토요일에 26.8%로 가장 높았고, 금요일에 1.5%로 가장 낮았음

<그림 3-25> 인천광역시 요일별 단속 현황

(기간: 2009 ~ 2018)

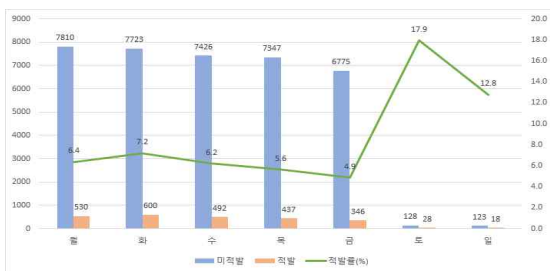


#### 5) 대구광역시 요일별 단속 현황

- 대구광역시는 토요일에 17.9%로 가장 높았고, 금요일에 4.9%로 가장 낮았음

<그림 3-26> 대구광역시 요일별 단속현황

(기간: 2009 ~ 2018)



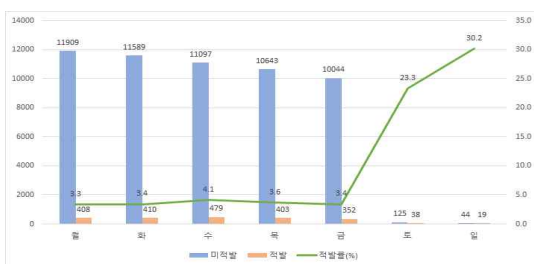
46

#### 6) 대전광역시 요일별 단속 현황

- 대전광역시는 일요일에 30.2%로 가장 높았고, 월요일에 3.3%로 가장 낮았음

<그림 3-27> 대전광역시 요일별 단속현황

(기간: 2009 ~ 2018)

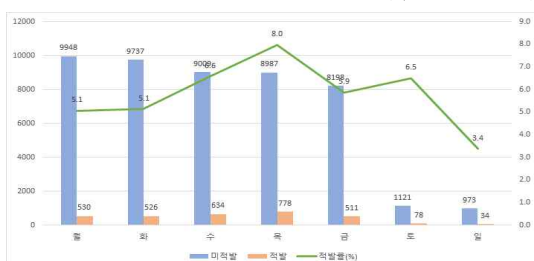


#### 7) 광주광역시 요일별 단속 현황

- 광주광역시는 목요일에 8.0%로 가장 높았고, 일요일에 3.4%로 가장 낮았음

<그림 3-28> 광주광역시 요일별 단속현황

(기간: 2009 ~ 2018)



47

## 사. 단속 현황

### 1) 단속 종류

- 원산지 단속은 매년 일정한 패턴으로 실시가 되는 정기단속, 특별한 이슈나 수입 급증 등에 따른 특별단속, 그리고 정기단속과 특별단속 시기를 제외한 평상시에 진행하는 상시단속 등으로 이루어져 있음
- 농관원은 <표 3-23>와 같이 매년 일정 시기별로 반복적인 정기단속을 진행하여 왔음
- 정기단속 시기 이외에는 수입이 급증하는 시기나 가격이 급등하는 경우와 같은 사회적 이슈가 있거나 부정유통 정보수집 등 기획조사가 필요한 경우(인삼류·한약재 등)에 특별단속을 하여 왔음
- 이외에도 민원 등이 접수되거나 평상시에는 상시단속을 해왔으며, 상시 조사 품목은 쇠고기·돼지고기·닭고기·쌀·배추김치·콩 등 6개 품목임

### 2) 주요 단속 내용

- 현장에서 이루어지는 단속방법은 아래와 같은 방식으로 이루어지고 있음
- 배추김치
  - 배추김치의 경우 관세청 통관 정보인 '수입농산물 유통이력 조회'를 이용하여 단속을 진행하며 관련 해당 정보는 농축산물유통관리시스템에서 연계조회 가능함. 해당 유통이력을 조회하여 중국산 배추김치를 구입한 업체를 대상으로 원산지 표시 단속 실시하며 단속 전 중국산 배추김치를 구입한 업체의 인터넷을 통해 블로그상의 원산지 표시사항을 확인하여 단속 실시함

48

- 학교급식 등 집단급식소 단속
  - 유치원 단속 시 공공형 어린이집이 아닌 민간 어린이집과 유치원을 대상으로 인원수 50인 이상의 대상업소를 원산지 단속 실시함. 학교급식 단속 시 해당 학교 홈페이지를 통해 일자별 급식메뉴를 확인하여 위반 개연성이 높은 품목이 제공되는 급식일에 원산지 점검실시를 하며 위반 개연성이 높은 품목(배추김치, 돼지불고기, 쇠고기 등) 위주로 단속·점검 실시

#### ○ 한약재 원산지 단속

- 한약재 중 수입 농산물 유통이력을 최종 소비자까지 등록 관리하는 품목 위주로 단속을 진행함. '수입농산물유통이력조회'를 통해 해당 품목을 구입한 업체를 위주로 단속 실시함

#### ○ 식육점 등

- 식육점 점검시 일반 식육점보다 위반 개연성이 높은 마트 내에 입점해 있는 식육점을 우선 단속 실시함(식약처가 운영하는 '식품정보활용시스템'에 업체명 '마트', 판매형태 '식육판매업'의 조건으로 검색하여 활용) → 시기별, 가격별 위반 개연성이 높은 품목(대패삼겹살, 냉장삼겹살, 양념육 등) 위주로 점검

#### ○ 육식점 단속

- 한식업체 중 위반 개연성이 높은 품목 '두루치기', '대패삼겹살', '양념돼지갈비'를 판매하는 업체 중 가격이 싼 업소를 우선적으로 단속 실시함

49

### 3) 정기 단속 월별 주요 내용

- 해마다 같은 패턴으로 정기 단속을 하고 있는데 그 내용을 정리하면 다음과 같음

<표 3-23> 정기단속 월별 주요 내용

월별	주요 내용
1~2월	○ 설 대비 제수·선물용품 원산지 표시단속 - 기간: '19.1.7.~2.1.(26일간) - 품목: 육류, 과일류, 나물류, 갈비·정육세트, 한과류, 버섯류 등
2월	○ 대보름 성수 농식품 원산지 표시단속 - 기간: '19.2.11.~2.18.(8일간) - 품목: 땅콩, 잣, 호두, 밤, 오곡, 고사리, 도라지, 버섯 등
3월	○ 학교급식업체 원산지 표시단속 - 기간: '19.3.11.~3.22.(12일간) - 품목: 당근, 마늘, 양파, 축산물 등
4월	○ 행락철 돼지고기 및 외국산 배추김치 원산지 표시단속 - 기간: '19.4.1.~4.10.(10일간) - 품목: 배추김치, 고춧가루, 마늘 등
5월	○ 수입 화훼류 원산지 표시단속 - 기간: '19.5.1.~5.15.(15일간) - 품목: 카네이션, 국화, 장미, 안개꽃, 백합, 산세베리아 등
7~8월	○ 휴가철 축산물 원산지 표시단속 - 기간: '19.7.15.~8.14.(31일간) - 품목: 쇠고기, 돼지고기, 닭고기, 오리고기, 염소고기
8~9월	○ 추석 대비 제수 및 선물용품 원산지 표시단속 - 기간: '19.8.19.~9.11.(24일간) - 품목: 육류, 과일류, 나물류, 갈비·정육세트, 한과류, 버섯류 등
11~12월	○ 김장철 배추김치 및 양념류 원산지 표시단속 - 기간: '19.11.4.~12.13.(40일간) - 품목: 배추김치, 고춧가루, 마늘, 생강, 당근 등

#### ○ 설 대비 제수용 농식품 단속

- 단속 대상 업체: 제수·선물용 농식품 제조·가공업체, 통신판매업체, 전통식품 제조업체, 주류제조업체, 전통시장, 도·소매시장, 양곡 판매상 등
- 주요 단속품목
  - 제수용: 육류(소·돼지·닭), 과일류(사과·배 등), 쌀, 나물류 등
  - 선물용: 갈비 세트, 송이버섯, 전통식품, 건강식품 등
  - 인삼류: 인삼류(미검사품 포함) 및 인삼 가공품 등
  - 기 타: 주류(탁주, 약주 등)
- \* 통신판매업체도 병행 단속(공통)

#### ○ 대보름 성수 농식품 원산지 단속

- 단속 대상 업체: 백화점, 할인매장, 농협 매장, 도매시장, 전통시장, 한약재 판매상 등(양곡도·소매상, 건과류 판매상, 참기름 제조·판매업체 등)
- 주요 단속품목
  - 양곡: 미곡류, 두류, 조, 수수쌀, 율무쌀, 기장쌀, 메밀 등
  - 건과류: 땅콩, 호두, 밤, 잣 등
  - 나물류: 도라지, 고사리, 취나물, 호박나물, 무말랭이, 토란대, 연근, 우영, 표고버섯, 콩나물, 숙주나물 등
  - 농산 가공품: 참기름, 들기름, 두부류, 목류 등

#### ○ 화훼류 원산지 표시 일제 단속

- 단속 대상 업체: 화훼공판장, 꽃 도·소매상(화원), 화환 제조·판매업체, 통신판매업체, 장례식장, 대형마트, 편의점 등
- 주요 단속품목: 카네이션, 국화, 장미 등 국산·외국산 절화류 위주 [국내산 절화류 원산지 대상품목(11): 국화, 카네이션, 장미, 백합, 글라디올러스, 튜립, 거베라, 아이리스, 프리지아, 칼라, 안개꽃]

#### ○ 휴가철 축산물 원산지 단속

- 단속 대상 업체: 축산물수입업체, 식육가공업체, 식육도·소매점, 전



- 통시장, 관광지 주변 음식점과 정육식당, 행사장 주변 푸드트럭 등
- 주요 단속품목: 쇠고기, 돼지고기, 양·염소고기, 닭고기, 오리고기, 식육부산물, 식육가공품 등

○ 추석 대비 제수용 농식품 단속: 설 대비 제수용 농식품 단속과 동일

○ 김장철 배추김치 등 원산지 단속

- 단속 대상 업체
  - 배추김치와 냉동고추·고춧가루 수입·유통업체, 김치 및 고춧가루 제조업체, 도·소매시장 식재료 취급업체, 대형 급식업체 및 일반음식점, 통신판매(홈쇼핑, 인터넷), 어린이집 및 학교급식 납품업체, 외국산 콩 취급업체
- 주요 단속품목
  - 배추김치, 냉동고추, 고춧가루, 마늘, 양파 등 조미채소, 콩, 콩나물, 콩비지, 콩국수, 두부(콩)

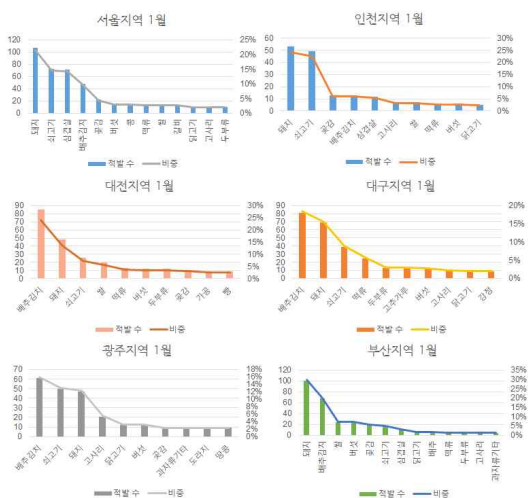
#### 4) 1~2월 정기 단속 현황(설)

○ 설 대비 제수·선물용품 원산지 표시단속을 주 내용으로 하여 설 전에 일정기간 동안 실시(2019년의 경우 1월 7일부터 2월 1일까지 26일간)하며 대상품목은 육류, 과일류, 나물류, 갈비·정육 세트, 한과류, 버섯류, 화훼류 등임

○ 2009년부터 2018년까지의 설 정기단속 현황을 분석한 결과 서울과 인천은 돼지고기와 쇠고기가 가장 많이 단속되었고 대구와 대전 그리고 부산은 배추김치와 돼지고기가 가장 많이 단속됨. 광주주는 배추김치와 쇠고기가 가장 많이 단속됨

<그림 3-29> 1월 정기 단속 현황

(기간: 2009 ~ 2018)



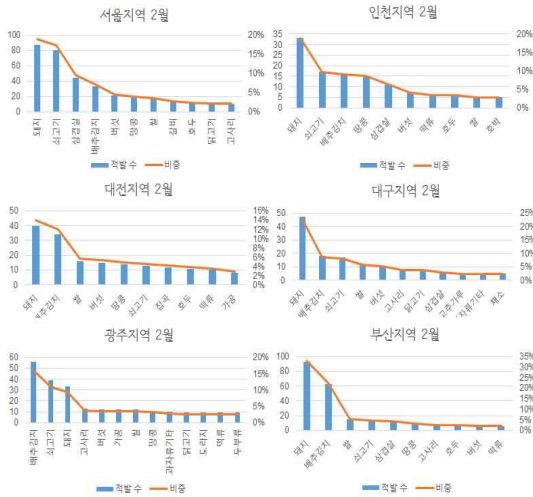
## 5) 2월 정기 단속 현황(대보름)

- 대보름 성수 농식품 원산지 표시 단속을 주 내용으로 하여 대보름 전에 일정기간 실시(2019년에는 2월 11일부터 2월 18일까지 8일간) 하며 대상품목은 땅콩, 잣, 호두, 밤, 오곡, 고사리, 도라지, 버섯 등임

- 주요 대상품목이 대보름 성수 농식품이나 각 지역별로 가장 많이 단속되는 품목은 돼지고기와 쇠고기, 배추김치로 나타남

<그림 3-30> 2월 정기 단속 현황

(기간: 2009 ~ 2018)



54

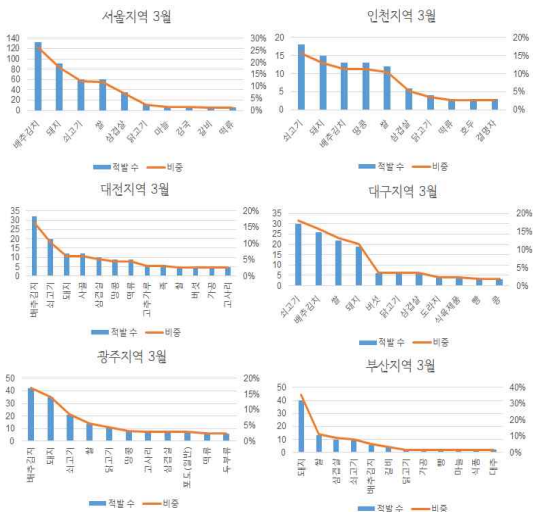
## 6) 3월 정기 단속 현황(학교 급식업체)

- 학교급식업체 원산지 표시단속을 주 내용으로 하여 3월 중에 실시 (2019년은 3월 11일부터 2019년 3월 22일까지 12일간)하며 대상품목은 당근, 마늘, 양파, 축산물 등임

- 지역별로 마늘과 떡류 등 급식 관련 식품이 전월에 비해 많이 단속되었으나 배추김치, 돼지고기, 쇠고기 등이 가장 많이 단속됨

<그림 3-31> 3월 정기 단속 현황

(기간: 2009 ~ 2018)



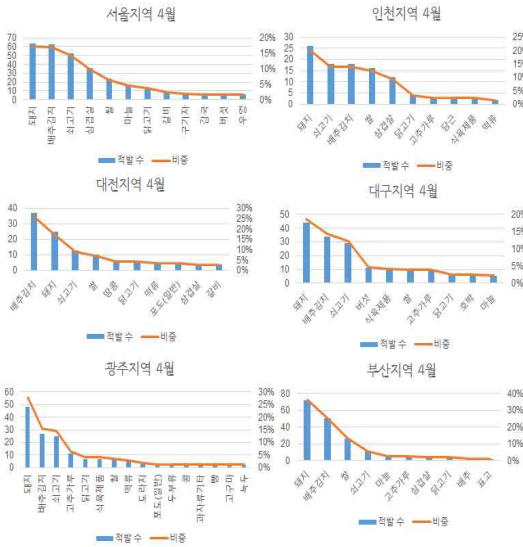
55

## 7) 4월 정기 단속 현황(행락철)

- 행락철 돼지고기 및 외국인 배추김치 원산지 표시 단속을 주 내용으로 하여 4월 중에 실시(2019년은 4월 1일부터 4월 10일까지 10일간)하며 대상품목은 배추김치, 고춧가루, 마늘 등임
- 고춧가루와 마늘이 전월에 비해 많이 단속되었으나 주로 돼지고기와 배추김치, 쇠고기가 가장 많이 단속됨
- 광주의 경우 타지역에 비해 고춧가루의 비율이 높게 나타남

<그림 3-32> 4월 정기 단속 현황

(기간: 2009 ~ 2018)



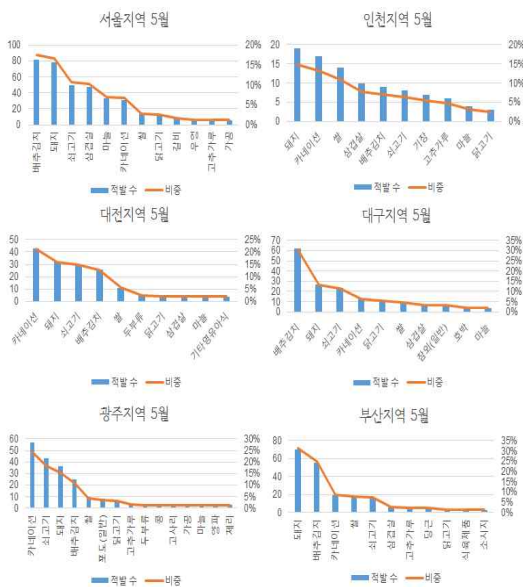
56

## 8) 5월 정기 단속 현황(수입 화훼류)

- 수입 화훼류 원산지 표시 단속을 주 내용으로 하며 5월 초에 실시 (2019년은 5월 1일부터 15일까지 15일간)하며 대상품목은 카네이션, 국화, 장미, 안개꽃, 백합, 산세베리아 등임
- 가정의 달 영향으로 화훼류 중 특히 카네이션의 단속이 높게 나타남

<그림 3-33> 5월 정기 단속 현황

(기간: 2009 ~ 2018)

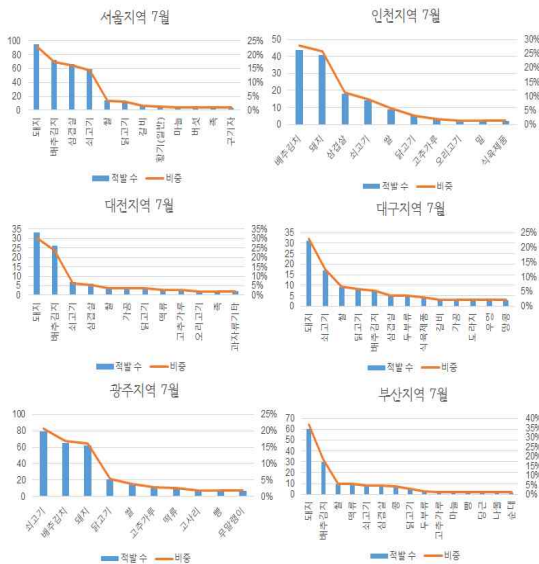


57

9) 7~8월 정기 단속 현황(휴가철)

- 휴가철 축산물 원산지 표시 단속을 주 내용으로 하여 7월과 8월에 걸쳐 실시(2019년은 7월 15일부터 8월 14일까지 31일간)하며 대상품목은 쇠고기, 돼지고기, 닭고기, 오리고기, 염소고기 등임
- 돼지고기와 배추김치 등이 주요 단속품목이고 특히 인천과 광주의 경우 전월 단속에 비해 배추김치와 돼지고기의 단속이 대폭 증가

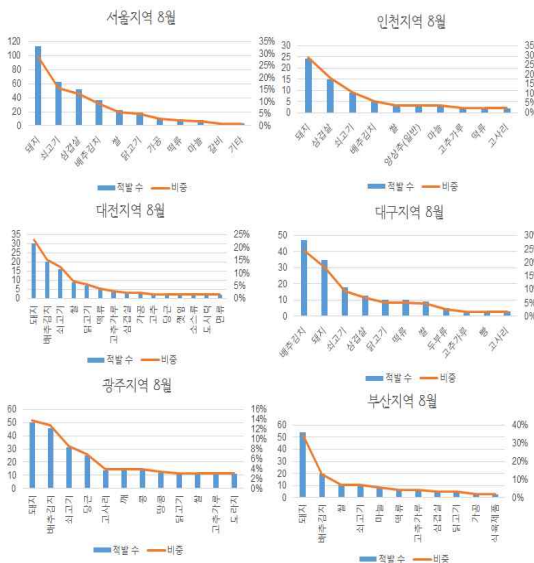
<그림 3-34> 7월 정기 단속 현황 (기간: 2009 ~ 2018)



10) 8~9월 정기 단속 현황(추석)

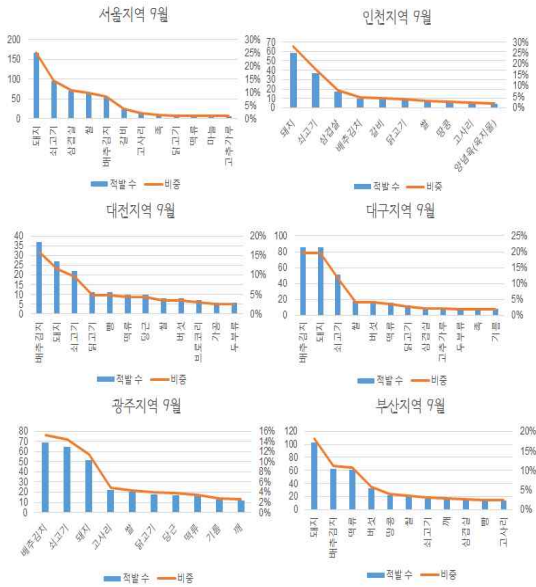
- 추석 대비 계수 및 선물용품 원산지 단속을 주 내용으로 하여 추석 전에 실시(2019년은 8월 19일부터 9월 11일까지 24일간)하며 대상품목은 육류, 고사리, 대추, 계수용, 선물용품 및 화훼류 등임
- 8월과 9월 양 월간 전체적인 단속이 증가하는데 특히 9월은 8월과 비교하여서도 단속이 크게 증가함

<그림 3-35> 8월 정기 단속 현황  
(기간: 2009 ~ 2018)



<그림 3-36> 9월 정기 단속 현황

(기간: 2009 ~ 2018)



60

#### 11) 11~12월 정기 단속 현황(김장철)

- 김장철 배추김치 및 양념류 원산지 표시단속을 주 내용으로 하여 11월 초부터 12월까지 실시(2019년은 11월 4일부터 12월 13일까지 40일간)하며, 대상품목은 배추김치, 고춧가루, 마늘, 생강, 당근 등임
- 시기적으로 김장철로 인해 전 지역에서 배추김치의 단속이 가장 많이 이루어짐

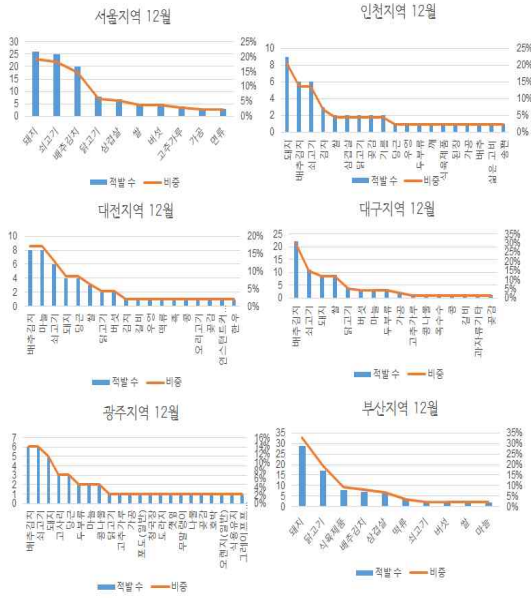
<그림 3-37> 11월 정기 단속 현황

(기간: 2009 ~ 2018)



61

<그림 3-38> 12월 정기 단속 현황  
(기간: 2009 ~ 2018)



## 아. 5대 품목별 단속 현황

### 1) 전국 5대 품목별 단속 현황

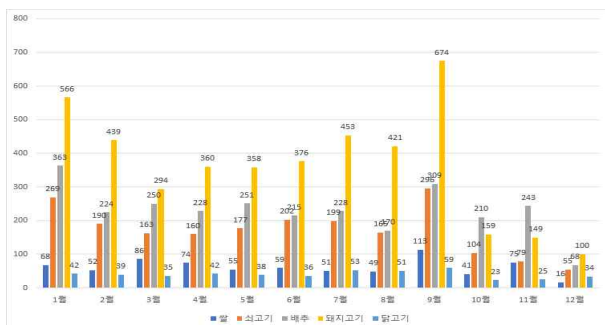
- 월별 품목별 단속 건수는 9월과 10월을 제외한 월에는 돼지고기가 가장 많이 단속되며 9월과 10월에는 배추김치의 단속 건수가 많음을 확인할 수 있음

<표 3-24> 전국 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	68	52	86	74	55	59	51	49	113	41	75	16
쇠고기	269	190	163	160	177	202	199	165	296	104	79	55
배추김치	363	224	250	228	251	215	228	170	309	210	243	68
돼지고기	566	439	294	360	358	376	453	421	674	159	149	100
닭고기	42	39	35	42	38	36	53	51	59	23	25	34

<그림 3-39> 전국 5대 품목별 단속 현황  
(기간: 2009 ~ 2018)



2) 광주광역시 5대 품목별 단속 현황

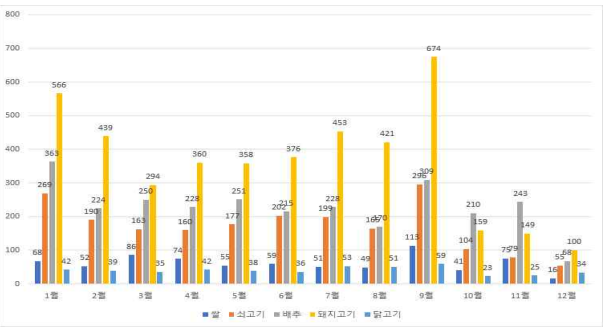
<표 3-25> 광주광역시 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	6	6	8	3	6	12	11	9	17	4	3	-
쇠고기	48	39	23	25	43	48	81	35	65	10	7	6
배추김치	67	59	40	25	24	32	52	42	68	31	29	6
돼지고기	53	45	39	50	39	69	65	53	65	14	18	6
닭고기	12	8	10	7	7	12	20	10	16	2	10	1

<그림 3-40> 광주광역시 5대 품목별 단속 현황

(기간: 2009 ~ 2018)



3) 대구광역시 5대 품목별 단속 현황

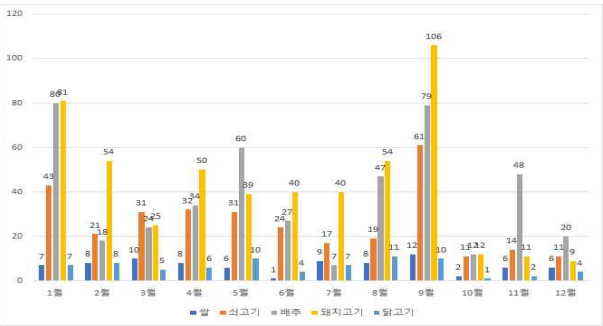
<표 3-26> 대구광역시 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	7	8	10	8	6	1	9	8	12	2	6	6
쇠고기	43	21	31	32	31	24	17	19	61	11	14	11
배추김치	80	18	24	34	60	27	7	47	79	12	48	20
돼지고기	81	54	25	50	39	40	40	54	106	12	11	9
닭고기	7	8	5	6	10	4	7	11	10	1	2	4

<그림 3-41> 대구광역시 5대 품목별 단속 현황

(기간: 2009 ~ 2018)



#### 4) 대전광역시 5대 품목별 단속 현황

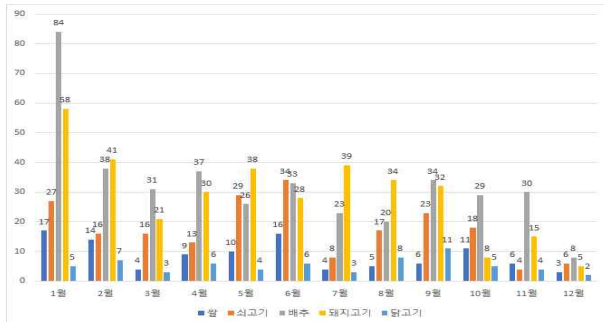
<표 3-27> 대전광역시 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	17	14	4	9	10	16	4	5	6	11	6	3
쇠고기	27	16	16	13	29	34	8	17	23	18	4	6
배추김치	84	38	31	37	26	33	23	20	34	29	30	8
돼지고기	58	41	21	30	38	28	39	34	32	8	15	5
닭고기	5	7	3	6	4	6	3	8	11	5	4	2

<그림 3-42> 대전광역시 5대 품목별 단속 현황

(기간: 2009 ~ 2018)



#### 5) 부산광역시 5대 품목별 단속 현황

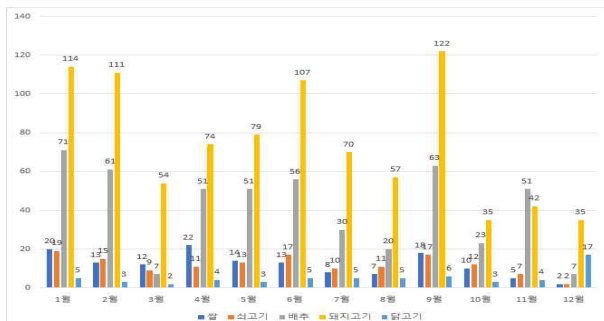
<표 3-28> 부산광역시 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	20	13	12	22	14	13	8	7	18	10	5	2
쇠고기	19	15	9	11	13	17	10	11	17	12	7	2
배추김치	71	61	7	51	51	56	30	20	63	23	51	7
돼지고기	114	111	54	74	79	107	70	57	122	35	42	35
닭고기	5	3	2	4	3	5	5	5	6	3	4	17

<그림 3-43> 부산광역시 5대 품목별 단속 현황

(기간: 2009 ~ 2018)





6) 서울특별시 5대 품목별 단속 현황

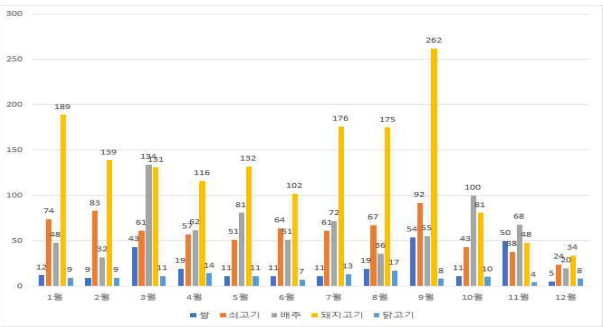
<표 3-29> 서울특별시 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	12	9	43	19	11	11	11	19	54	11	50	5
쇠고기	74	83	61	57	51	64	61	67	92	43	38	24
배추김치	48	32	134	62	81	51	72	36	55	100	68	20
돼지고기	189	139	131	116	132	102	176	175	262	81	48	34
닭고기	9	9	11	14	11	7	13	17	8	10	4	8

<그림 3-44> 서울특별시 5대 품목별 단속 현황

(기간: 2009 ~ 2018)



7) 인천광역시 5대 품목별 단속 현황

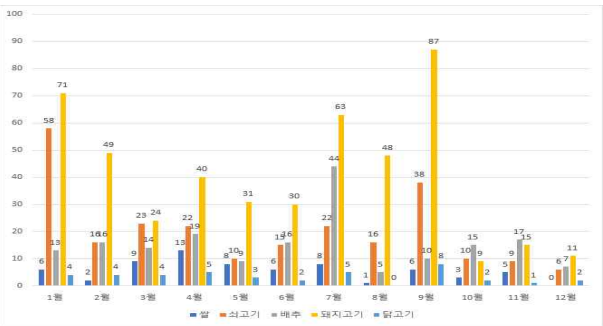
<표 3-30> 인천광역시 5대 품목별 단속 현황

단위: 건, 기간: 2009 ~ 2018

위반내역	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
쌀	6	2	9	13	8	6	8	1	6	3	5	-
쇠고기	58	16	23	22	10	15	22	16	38	10	9	6
배추김치	13	16	14	19	9	16	44	5	10	15	17	7
돼지고기	71	49	24	40	31	30	63	48	87	9	15	11
닭고기	4	4	4	5	3	2	5	-	8	2	1	2

<그림 3-45> 인천광역시 5대 품목별 단속 현황

(기간: 2009 ~ 2018)



## 자. 위반 금액별 단속 현황

### 1) 지역별 위반 금액 비중

- 전체 위반 금액의 단속빈도는 5만원 미만이 4,391(26.5%)건으로 가장 높으며 10만원 이상 100만원 미만 4,127(24.9%)건, 100만원 이상 1000만원 미만은 3,406(20.5%)건 순으로 나타남

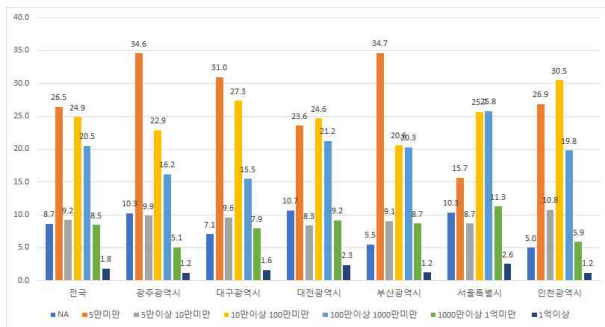
<표 3-31> 지역별 위반 금액 비중

단위: %, 기간: 2009 ~ 2018

지역	NA	5만원미만	5만원이상 10만원미만	10만원이상 100만원미만	100만원이상 1000만원미만	1000만원이상 1억원미만	1억원이상
전국	8.7	26.5	9.2	24.9	20.5	8.5	1.8
광주광역시	10.3	34.6	9.9	22.9	16.2	5.1	1.2
대구광역시	7.1	31.0	9.6	27.3	15.5	7.9	1.6
대전광역시	10.7	23.6	8.3	24.6	21.2	9.2	2.4
부산광역시	5.5	34.7	9.1	20.6	20.3	8.7	1.2
서울특별시	10.4	15.7	8.7	25.7	25.8	11.3	2.6
인천광역시	5.0	26.9	10.8	30.5	19.8	5.9	1.2

<그림 3-46> 지역별 위반 금액 비중

(기간: 2009 ~ 2018)



## 차. 업체별 단속 현황

### 1) 업체 축소 후 분석 진행

- 36개의 업체를 7개(일반음식점, 건강기능식품, 식품제조가공, 식품판매, 식육점, 급식, 기타)로 축소해서 진행

<그림 3-47> 업체 축소 현황

일반음식점	식품판매	식육점	기타
일반음식점	건어물상	축산물유통 전문판매업	기타
휴게음식점	농협	식육판매업	DSC
	대형유통업체(양곡)	식육포장처리업	기타(가공업체)(양곡)
건강기능식품	도매상, 소매업체(양곡)	축협	노점상
건강기능식품판매업	백화점, 소매상(양곡)		통신판매업체
한약상	소분업자	급식	RPC
	수입상, 슈퍼	미학교급식업체	기타(식품제조가공)(양곡)
식품제조가공	식품유통업	급식자재납품업체	종묘상
가공업체	양곡상, 계곡점영업	위탁급식소	
생산업체	청과상, 할인매장	집단급식소	
생산자기타	식품판매(양곡)	학교급식업체	
임도정공장	생산자단체매장(양곡)	조달청급식	
	식품점		

## 2) 업태별 단속률: 전국

- 업태별 단속률은 일반음식점이 48.7%로 가장 높고, 식육점 14.7%, 기타 12.6%, 식품판매업 12.5%, 식품제조가공 8.9% 순으로 나타남

<표 3-32> 전국 업태별 단속률

단위: 건, 기간: 2009 ~ 2018

구분	합계	일반음식점	식품판매	식육점	식품제조가공	기타	건강기능식품	급식
적발	16,769	8,164	2,102	2,473	1,492	2,109	192	237
단속률(%)	100	48.7	12.5	14.8	8.9	12.6	1.1	1.4

<그림 3-48> 전국 업태별 단속률

(기간: 2009 ~ 2018)



72

## 3) 업태별 추이 분석: 전국

- 전국에서 일반음식점(251,000개소)이 업태 중 가장 많은 분포를 차지하고 있으며 식품판매(82,953개소), 식육점(72,718개소) 순으로 나타남
- 업태별 위반율에서는 기타를 제외하고 식품제조가공 업태가 가장 높은 위반율(7.7%)을 보임

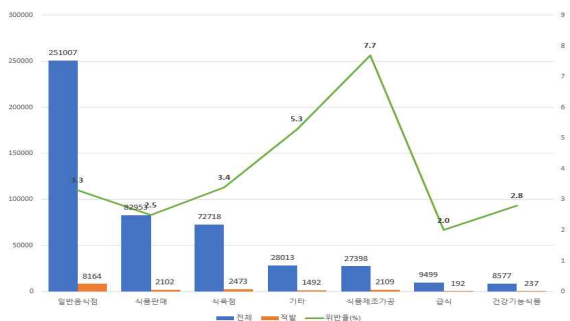
<표 3-33> 전국 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식품판매	식육점	식품제조가공	기타	건강기능식품	급식
전체	251,007	82,953	72,718	28,013	27,398	9,499	8,577
적발	8,164	2,102	2,473	1,492	2,109	192	237
위반율(%)	3.2	2.5	3.4	5.3	7.7	2.0	2.8

<그림 3-49> 전국 업태 분포

(기간: 2009 ~ 2018)



1) 위반율 = 적발 업태/업태 전체

73

4) 업태별 추이 분석: 서울특별시

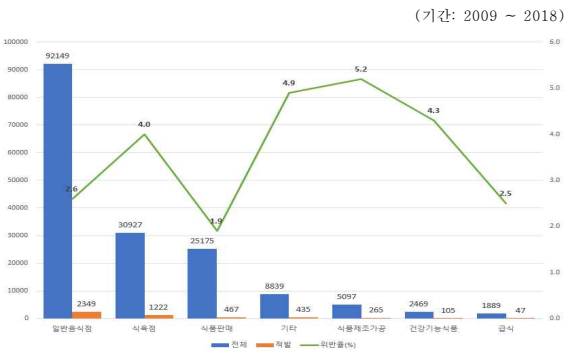
- 서울 지역에서는 일반음식점(92,149개소)이 업태 중 가장 많은 분포를 차지
- 서울 지역에서는 식품제조가공 업태가 가장 높은 위반율을 보임

<표 3-34> 서울 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식육점	식품판매	기타	식품제조가공	건강기능식품	급식
전체	92,149	30,927	25,175	8,839	5,097	2,469	1,889
적발	2,349	1,222	467	435	265	105	47
위반율(%)	2.6	4.0	1.9	4.9	5.2	4.3	2.5

<그림 3-50> 서울 업태 분포



5) 업태별 추이 분석: 인천광역시

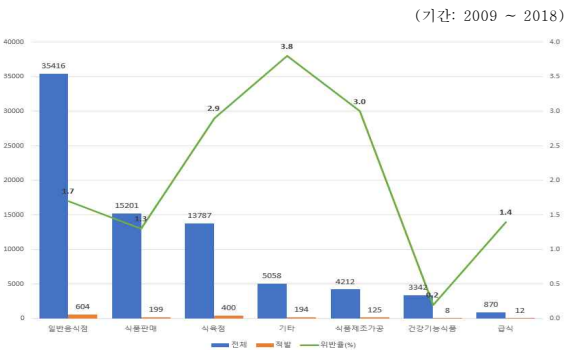
- 인천 지역에서는 일반음식점이 업태 중 가장 많은 분포를 차지
- 인천 지역에서는 기타를 제외하면 식품제조가공과 식육점 업태가 가장 높은 위반율을 보임

<표 3-35> 인천 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식품판매	식육점	기타	식품제조가공	건강기능식품	급식
전체	35,416	15,201	13,787	5,058	4,212	3,342	870
적발	604	199	400	194	125	8	12
위반율(%)	1.7	1.3	2.9	3.8	3.0	0.2	1.4

<그림 3-51> 인천 업태 분포



6) 업태별 추이 분석: 부산광역시

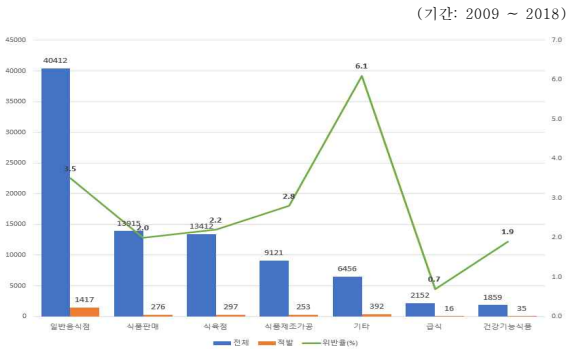
- 부산 지역에서는 일반음식점이 업태 중 가장 많은 분포를 차지
- 부산 지역에서는 기타를 제외하면 일반음식점 업태가 가장 높은 위반율을 보임

<표 3-36> 부산 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식품판매	식육점	식품제조가공	기타	급식	건강기능식품
전체	40,412	13,915	13,412	9,121	6,456	2,152	1,859
적발	1,417	276	297	253	392	16	35
위반율(%)	3.5	2.0	2.2	2.8	6.1	0.7	1.9

<그림 3-52> 부산 업태 분포



7) 업태별 추이 분석: 대전광역시

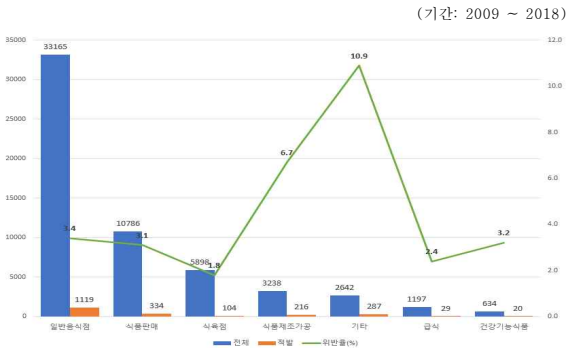
- 대전 지역에서는 일반음식점이 업태 중 가장 많은 분포를 차지
- 대전 지역에서는 기타를 제외하면 식품제조가공 업태가 가장 높은 위반율을 보임

<표 3-37> 대전 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식품판매	식육점	식품제조가공	기타	급식	건강기능식품
전체	33,165	10,786	5,898	3,238	2,642	1,197	634
적발	1,119	334	104	216	287	29	20
위반율(%)	3.4	3.1	1.8	6.7	10.9	2.4	3.2

<그림 3-53> 대전 업태 분포



8) 업태별 추이 분석: 대구광역시

- 대구 지역에서는 일반음식점이 업태 중 가장 많은 분포를 차지
- 대구 지역에서는 기타를 제외하면 식품제조가공 업태가 가장 높은 위반율을 보임

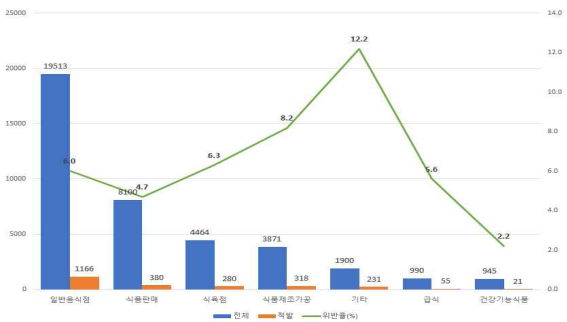
<표 3-38> 대구 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식품판매	식육점	식품제조가공	기타	급식	건강기능식품
전체	19,513	8,100	4,464	3,871	1,900	990	945
적발	1,166	380	280	318	231	55	21
위반율(%)	6.0	4.7	6.3	8.2	12.2	5.6	2.2

<그림 3-54> 대구 업태 분포

(기간: 2009 ~ 2018)



9) 업태별 추이 분석: 광주광역시

- 광주 지역에서는 일반음식점이 업태 중 가장 많은 분포를 차지
- 광주 지역에서는 기타를 제외하면 식품제조가공 업태가 가장 높은 위반율을 보임

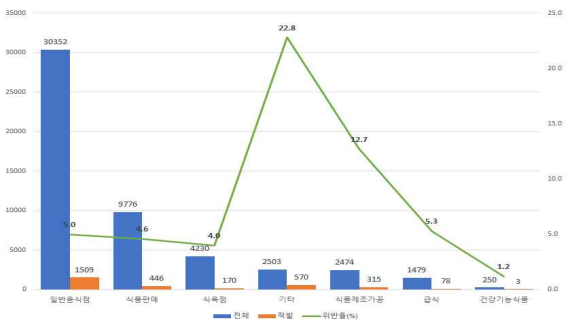
<표 3-39> 광주 업태 분포

단위: 건, 기간: 2009 ~ 2018

구분	일반음식점	식품판매	식육점	기타	식품제조가공	급식	건강기능식품
전체	30,352	9,776	4,230	2,503	2,474	1,479	250
적발	1,509	446	170	570	315	78	3
위반율(%)	5.0	4.6	4.0	22.8	12.7	5.3	1.2

<그림 3-55> 광주 업태 분포

(기간: 2009 ~ 2018)



카. 재워반 경향성 분석

1) 데이터 전처리 과정

- [실체분자 + 실체분자주민번호 + 연도/월]를 key값으로 사용하여 업소를 구분
- 2009년부터 2018년까지 적발된 업소들을 연도별, 월별로 집계하여 120개월 중 2개월 이상 적발된 업소를 재적발 업소로 지정

2) 지역별 재워반 분석

- 인천광역시 15.6%로 가장 높고, 대전광역시 8.6%로 가장 낮은 재적발률을 보임

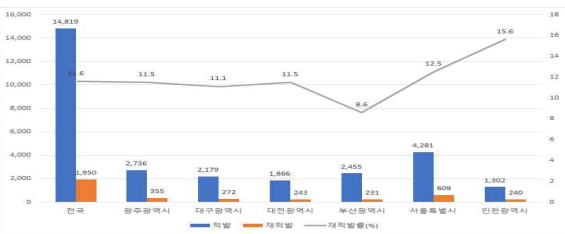
<표 3-40> 지역별 재적발률

단위: 건, 기간: 2009 ~ 2018

시도	적발	재적발	재적발률(%)
전국	14,819	1,950	11.6
광주광역시	2,736	355	11.5
대구광역시	2,179	272	11.1
대전광역시	1,866	243	11.5
부산광역시	2,455	231	8.6
서울특별시	4,281	609	12.5
인천광역시	1,302	240	15.6

<그림 3-56> 도시별 재워반율

(기간: 2009 ~ 2018)



3) 업주 연령별 재적발률

- 결측값을 제외하고 재적발률은 전국에서 40대가 32.8%로 가장 높은 수준으로 나타났으며 50세 이상이 22.5%로 두 번째로 높음
- 대구, 대전, 서울의 경우 40대에서 높은 재적발률을 보이며 광주는 50대에서 부산, 인천은 30대에서 가장 높은 재적발률을 보임

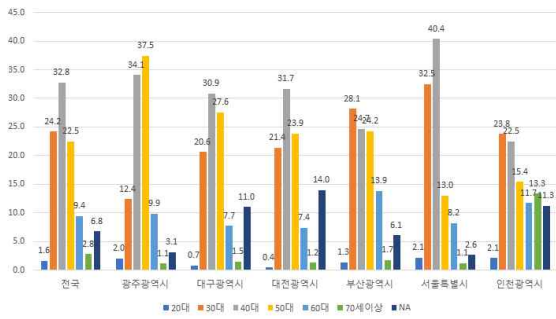
<표 3-41> 업주 연령에 따른 재워반율

단위: 건, 기간: 2009 ~ 2018

시도	구분	20대	30대	40대	50대	60대	70세이상	NA
전국	적발	308	2,434	4,833	4,866	1,792	296	290
	재적발	31	472	639	438	184	54	132
	재적발률(%)	1.6	24.2	32.8	22.5	9.4	2.8	6.8
광주광역시	적발	63	424	990	888	292	45	34
	재적발	7	44	121	133	35	4	11
	재적발률(%)	2.0	12.4	34.1	37.5	9.9	1.1	3.1
대구광역시	적발	42	358	728	714	246	25	66
	재적발	2	56	84	75	21	4	30
	재적발률(%)	0.7	20.6	30.9	27.6	7.7	1.5	11.0
대전광역시	적발	51	272	636	684	182	24	17
	재적발	1	52	77	58	18	3	34
	재적발률(%)	0.4	21.4	31.7	23.9	7.4	1.2	14.0
부산광역시	적발	40	334	612	916	432	73	48
	재적발	3	65	57	56	32	4	14
	재적발률(%)	1.3	28.1	24.7	24.2	13.9	1.7	6.1
서울특별시	적발	79	799	1,436	1,287	520	96	64
	재적발	13	198	246	79	50	7	16
	재적발률(%)	2.1	32.5	40.4	13.0	8.2	1.1	2.6
인천광역시	적발	33	247	431	377	120	33	61
	재적발	5	57	54	37	28	32	27
	재적발률(%)	2.1	23.8	22.5	15.4	11.7	13.3	11.3

<그림 3-57> 업주 연령별 재위반율

(기간: 2009 ~ 2018)



#### 4) 위반 물량별 제적발 분석

- 위반 물량별 제적발률은 전국에서 10kg미만의 물량에서 59.1%로 가장 높은 제적발률을 보임
- 10~100kg 미만인 20.2%로 10kg 미만(59.1%)과 합계를 하면 100kg 미만이 79.3%로 재위반의 대다수를 차지하며 물량이 작을수록 재위반 경향이 높음

<표 3-42> 위반 물량별 제적발률

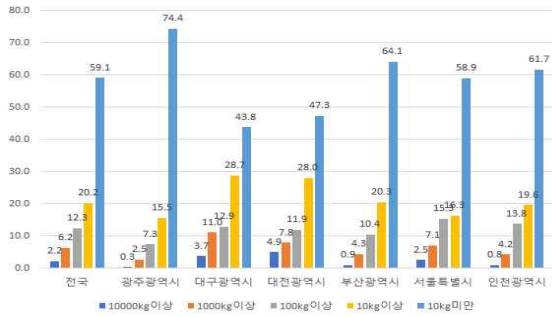
단위: 건, 기간: 2009 ~ 2018

시도	구분	10000kg이상	1000kg이상	100kg이상	10kg이상	10kg미만
전국	적발	198	850	2,638	3,532	7,601
	재적발	42	121	240	394	1,153
	재적발률(%)	2.2	6.2	12.3	20.2	59.1
광주광역시	적발	16	124	463	575	1,558
	재적발	1	9	26	55	264
	재적발률(%)	0.3	2.5	7.3	15.5	74.4
대구광역시	적발	50	125	332	682	990
	재적발	10	30	35	78	119
	재적발률(%)	3.7	11.0	12.9	28.7	43.8
대전광역시	적발	25	131	382	391	937
	재적발	12	19	29	68	115
	재적발률(%)	4.9	7.8	11.9	28.0	47.3
부산광역시	적발	25	83	347	624	1,376
	재적발	2	10	24	47	148
	재적발률(%)	0.9	4.3	10.4	20.3	64.1
서울특별시	적발	57	313	864	896	2,151
	재적발	15	43	93	99	359
	재적발률(%)	2.5	7.1	15.3	16.3	58.9
인천광역시	적발	25	74	250	364	589
	재적발	2	10	33	47	148
	재적발률(%)	0.8	4.2	13.8	19.6	61.7



<그림 3-58> 전국 위반 물량별 제위반율

(기간: 2009 ~ 2018)



##### 5) 위반 금액별 제위반 분석

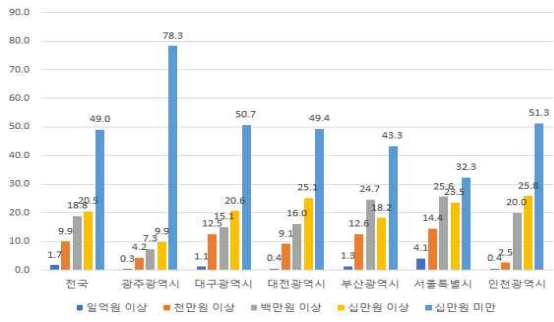
- 위반 금액별 제적발률은 전국에서 10만원 미만의 위반 금액에서 49.0%로 가장 높은 제적발률을 보임

<표 3-43> 전국 위반 금액별 제위반율

단위: 건, 기간: 2009 ~ 2018

시도	구분	일억원 이상	천만원 이상	백만원 이상	십만원 이상	십만원 미만
전국	적발	228	1,135	2,873	3,148	7,435
	재적발	34	194	367	399	956
	재적발률(%)	1.7	9.9	18.8	20.5	49.0
광주광역시	적발	28	127	450	567	1,564
	재적발	1	15	26	35	278
	재적발률(%)	0.3	4.2	7.3	9.9	78.3
대구광역시	적발	27	138	299	475	1,240
	재적발	3	34	41	56	138
	재적발률(%)	1.1	12.5	15.1	20.6	50.7
대전광역시	적발	36	156	372	394	908
	재적발	1	22	39	61	120
	재적발률(%)	0.4	9.1	16.0	25.1	49.4
부산광역시	적발	27	193	453	440	1,342
	재적발	3	29	57	42	100
	재적발률(%)	1.3	12.6	24.7	18.2	43.3
서울특별시	적발	94	442	1,057	964	1,724
	재적발	25	88	156	143	197
	재적발률(%)	4.1	14.4	25.6	23.5	32.3
인천광역시	적발	16	79	242	308	657
	재적발	1	6	48	62	123
	재적발률(%)	0.4	2.5	20.0	25.8	51.3

<그림 3-59> 전국 위반 금액별 제위반율  
(기간: 2009 ~ 2018)



## IV 활용데이터 설명 및 현황

### 1 업체 및 지역

- 본 연구에서는 업체추천과 그리드 단위의 지역추천 두 가지로 나누어 진행하였으며 사용된 데이터와 변수도 차이가 있음

#### 가. 업체추천 데이터

- 업체추천의 경우 서울을 포함한 주요 도시의 자료를 모두 포함하고 있어 전국 단위의 자료를 사용하거나 각 시/구 단위로 수집한 데이터를 취합하여 활용함

<표 4-1> 활용 데이터 목록(업체추천)

1	영업 인허가 대장	9	유흥업소, 단란주점, 노래연습장 정보
2	단속실적	10	전통시장 표준데이터
3	어린이집 표준데이터	11	축산판매업 정보
4	초중등교육기관	12	전국 어린이 보호구역
5	유통전문판매업	13	전국 음식점
6	집단급식소	14	전국 병원
7	전국 대학 및 전문대학 DB정보	15	전국 식품제조가공업
8	위탁급식업체 정보	-	-

- 사용된 데이터의 출처와 내용 등은 <표 4-2> 등에 명기되어 있음

<표 4-2> 업체추천 분석 데이터 출처 및 설명

자료명	출 처	설 명	비 고
영업인허가 대장	국립농산물품질 관리원	- 농관원 내부자료	총 1,577,014
조사내역			조사업체: 470,063
단속실적			단속업체: 17,042
어린이집 표준 데이터	공공데이터 포털	- 전국 어린이집 정보(어린이집 유형, 정원 수 등)	2018년 5월 15일 기준
전국교육기관	공공데이터 포털	- 전국 유초중등학교위치 정보(학교명, 학교 위치 등)	지자체 개별 데이터 취합
유통 전문 판매업	지방행정인허가 데이터개방	- 직접 제조하지 않고 식가공류 제공을 위 해 판매업소에 유통하는 업소정보	지자체 개별 데이터 취합
전국 대학 및 전문대학DB 정보	공공데이터 포털	- 전국 대학 및 전문대학의 종류별 학교명 및 상태, 주소 등 관련 정보	2016년 3월 10일 기준
집단급식소 정보	지방행정인허가 데이터개방	- 병원, 학교, 어린이집 등 집단급식소를 운 영하는 업소 및 기관 정보	2018년 11월 12일 기준
위탁급식업체 정보	지방행정인허가 데이터개방	- 집단급식소를 설치 및 운영하는 업체와 계약에 따라 해당 집단급식소에서 음식류 를 조리하여 제공하는 업소정보	2018년 11월 12일 기준

<표 4-3> 업체추천 분석 데이터 출처 및 설명(계속)

자료명	출 처	설 명	비 고
유흥업소, 단란주점, 노래연습장 정보	지방행정인허가 데이터개방	- 서울시에 있는 식품접객업(일반음식점, 유흥 주점, 단란주점, 노래연습장) 현황 및 주 소 정보	2016년 12월 23일 기준
전국 전통시장 표준데이터	공공데이터 포털	- 전국 전통시장 정보(시장유형, 개설 주기 등)를 제공/서울 지역 정보만 추출	2018년 5월 15일 기준
축산판매업 정보	지방행정인허가 데이터개방	- 달걀, 유제품, 내장 또는 쇠고기, 돼지고기 와 같은 축산물을 판매하는 업소정보	지자체 개별 데이터 취합
어린이 보호구역	공공데이터 포털	- 전국 어린이보호구역 정보(어린이보호구역 장소, 보호구역내cctv설치정보 등)를 제공	
일반음식점	지방행정인허가 데이터개방	- 한식, 중식, 일식 등 음식류 를 조리 및 판매하며, 식사 와 함께 음주행위가 허용 되는 업소정보	
전국 병원	공공데이터 포털	- 병원, 종합병원, 치과병원 등 의사가 환자 를 진찰 및 진료하는데 필요한 설비가 갖 춰있으며, 30명 이상의 환자를 수용할 수 있는 시설을 갖춘 의료기관 정보	
식품제조 가공업	지방행정인허가 데이터개방	- PB제품, 도시락 등 식품을 제공 및 가공 하는 업소정보	

## 나. 지역추천 데이터

- 지역추천의 경우 서울을 대상으로 분석을 진행하였으며 대부분 서울시 통계 포털에서 데이터를 획득함
- 서울의 경우 타지역에 대비해 가용한 자료가 훨씬 많아 분석이 용이함

<표 4-4> 활용 데이터 목록(지역추천)

1	영업 인허가 대장	8	집단급식소 식품판매업체 정보
2	단속실적	9	서울 행정구역 면적
3	유흥업소, 단란주점, 노래연습장 정보	10	동별 소득세 데이터
4	어린이집 표준데이터	11	서울시 대학 및 전문대학 DB 정보
5	시장 표준데이터	12	서울시 동 공무원 통계
6	서울특별시 요양병원(병원급) 정보	13	원산지 표시 우수업체 현황
7	위탁 급식업체 정보	14	서울특별시 시장, 상점가(동별)면적

- 사용된 데이터의 출처와 내용 등은 <표 4-5> 등에 명기되어 있음

<표 4-5> 지역추천 분석 데이터 출처 및 설명

자료명	출처	설명	비고
영업인허가대장	농관원	- 농관원 내부자료	총 1,577,014
조사내역			조사업체: 470,063
단속실적			단속업체: 17,042
서울 유흥업소, 단란주점, 노래연습장 정보	서울특별시	- 서울시에 있는 식품접객업(일반음식점, 유흥주점, 단란주점, 노래연습장) 현황 및 주소 정보	2016년 12월 23일 기준
서울 어린이집 표준데이터	공공데이터 포털 (서울특별시 각 구청)	- 전국 어린이집 정보(어린이집 유형, 정원수 등)	2018년 5월 15일 기준
서울 전통시장 표준데이터	공공데이터 포털	- 전국 전통시장 정보(시장유형, 개설 주기 등)를 제공/서울 지역 정보만 추출	2018년 5월 15일 기준
서울특별시 요양병원(병원급) 정보	서울특별시	- 서울특별시에 소재하는 요양환자 30인 이상 수용할 수 있는 시설을 갖춘 의료기관 정보	2018년 11월 12일 기준
서울특별시 위탁급식업체 정보	서울특별시	- 서울특별시 위탁 급식업체 정보 업소명, 주소, 연락처, 위치정보 등을 제공	2018년 11월 12일 기준
서울특별시 집단급식소 식품판매업체 정보	서울특별시	- 서울특별시 위탁 급식업체 정보 업소명, 주소, 연락처, 위치정보 등을 제공	2018년 11월 12일 기준

자료명	출처	설명	비고
서울특별시 행정구역(동별) 면적	서울특별시	- 서울시 각 자치구 동별 행정구역의 면적 및 하위 행정구역 현황	2012년 2월 29일 기준으로 공개된 데이터 / 매년 12월 갱신
구청별 월별 소득세 징수실적	서울특별시	- 소득세, 정보공개청구로 서울시로부터 받은 자료	2009년 1월~2018년 12월 (3년 치)
서울시 대학 및 전문대학 DB정보	서울특별시	- 서울 시내 대학 및 전문대학의 종류별 학교명 및 상태, 주소 등 관련 정보	2016년 3월 10일 기준으로 수정된 데이터
서울시 동별 공무원 통계	서울특별시	- 서울시 동 공무원의 직류별, 직급별 정원(동별)현황	2017년 12월 26일 기준으로 공개된 데이터 / 매년 12월 갱신
국립 농산물 품질 관리원 원산지 표시 우수업체 현황	공공데이터 포털 (농관원)	- 국립농산물 품질 관리원에서 관리하는 원산지 표시 우수업체 현황 정보(승인번호, 승인일자, 업종구분, 업체명, 주소)	2019년 4월 18일 기준으로 수정된 데이터
서울특별시 시장, 상점가 (동별)면적	서울특별시	- 서울시 각 자치구 동별 등록, 인정시장, 상점가의 개소 수, 점포수, 면적 등	2017년 12월 26일 기준으로 공개된 데이터 / 매년 12월 갱신
식품제조 가공업	지방행정인허가 데이터개방	- PB제품, 도시락 등 식품을 제공 및 가공하는 업소정보	개별 파일별 갱신 시점이 다름

## 2 외부요인 분석 - 뉴스분석

- 외부 요인 분석을 하기 위해 뉴스 분석을 하였으며 뉴스 분석을 위해 먼저 신문기사를 수집하고 수집한 데이터를 기준으로 형태소 분석을 진행하였음
- 형태소 분석결과를 통해 키워드를 추출하고 추출한 키워드를 활용하여 원산지 표시에 영향을 줄 수 있는 요인들로 감성사전을 제작함
- 제작한 감성사전을 기준으로 각 기사별 감성점수를 부여하여 변수로 활용함

<그림 4-1> 뉴스 분석 진행도



### 1) 데이터 수집

- 분석을 위한 뉴스 기사는 크롤링 기법을 활용하여 중앙지 11개, 지역 종합지 28개, 주요 방송사 5개에서 2009년부터 2018년까지 돼지고기, 쇠고기, 닭고기, 배추김치, 쌀 등 총 6개 키워드로 29만5천개의 기사를 수집하였음

<표 4-6> 검색어 별 수집기사 건수

검색어	수집 기사(건)
쇠고기	47,928
돼지고기	45,062
닭고기	16,732
배추김치	53,892
쌀	85,860
원산지	45,763
총합	295,237

- 수집한 데이터를 기반으로 형태소 분석을 진행하였으며 형태소 분석 결과를 기준으로 감성사전을 제작하였음. 감성사전에서 원산지 표시 위반에 긍정적으로 작용하는 요소들은 긍정으로, 원산지 표시 위반에 부정적으로 작용할 수 있는 요소들은 부정으로 정의함
- 국내가격상승, 수입가격 하락, 국내물량 하락, 수입물량 상승을 긍정으로, 국내가격 하락, 수입가격 상승, 국내물량 상승, 수입물량 하락과 관련된 뉴스를 부정으로 정의함
- 앞서 도출한 긍·부정 감성사전을 중심으로 기사별 감성을 분석함. 기사가 긍정일 경우 1로, 부정일 경우 -1로 점수를 부여하여 이를 월별로 합산하여 해당 월의 품목별 뉴스 점수를 도출하였음

<그림 4-2> 감성점수 도출 예시

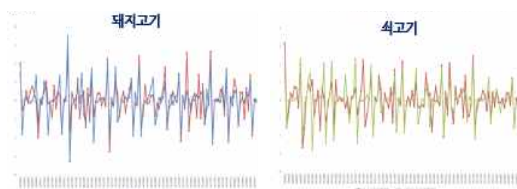
positive_score	negative_score	senti_score	new
2	-4	-2	-1
10	-10	0	0
0	0	0	0
14	-20	-6	-1
0	0	0	0
6	-2	4	1
0	0	0	0

- 뉴스기사의 감성분석 결과를 검정하기 위하여 실제 단속데이터의 월별 증감률과 감성점수의 증감률 간의 상관관계를 시간지연효과(lag)를 고려하여 분석하기 위해 ARIMA<sup>1)</sup> 모형을 사용해 검정함
- ARIMA 검정 결과 돼지고기가 76.4%, 쇠고기가 64.6%, 배추김치가 38.6%, 닭고기가 22.7%, 쌀이 25.5%의 상관관계를 보임(<표 4-7> 참조)
- 특히 돼지고기와 쇠고기는 상관관계도 매우 높은 편이며 lag도 존재하지 않아 이후 업체추천에서 유의미한 변수로 활용함

<표 4-7> ARIMA 분석결과

검색어	상관계수	시간차
쇠고기	0.646	-
돼지고기	0.764	-
닭고기	0.227	10
배추김치	0.386	10
쌀	0.254	-

<그림 4-3> 월별 뉴스 점수 증감률 품목별 단속 건수 증감률 비교 (돼지고기, 쇠고기)



1) ARIMA: 시계열 분석 기법의 한 종류로, 과거의 관측값과 오차를 사용해서 현재의 시계열 값을 설명하는 ARIMA 모형을 일반화 한 모형임

## V 알고리즘 설명

### 1 머신러닝을 활용한 단속 업체추천

#### 가. 모형 개요

- 머신러닝은 인공지능의 한 분야로 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야를 말함
- 머신러닝은 컴퓨터에게 사람이 직접 명시적으로 논리를 지시하지 않아도 데이터를 통해 컴퓨터가 '학습'을 하고 그것을 사용해 컴퓨터가 자동으로 문제를 해결하도록 하는 것임
- 머신러닝의 모형에는 크게 지도학습과 비지도학습이 있음. 지도학습은 종속변수<sup>1)</sup>와 독립변수<sup>2)</sup> 사이의 패턴을 학습하고 예측하는 것이고 비지도학습은 종속변수가 없고 데이터의 독립변수만 가지고 있으므로 독립변수의 패턴, 특성을 학습을 통해 발견함
- 본 연구에서 사용된 알고리즘은 모두 지도학습 모형임

#### 나. 머신러닝 모형 종류

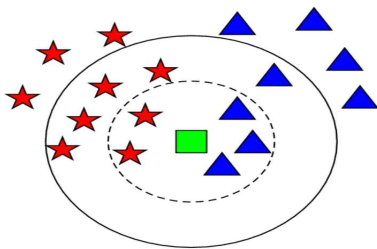
- 본 연구는 머신러닝 알고리즘 중에서도 대표적인 지도학습 분류 모형인 KNN, 로지스틱 회귀, 의사결정나무, 랜덤 포레스트, 그래디언트 부스팅, 에이다 부스팅, 다층 퍼셉트론, 서포트 벡터 머신 등 총 8개 모형으로 분석하여 가장 정확도가 높은 모형의 분석결과를 활용함

1) 종속변수: 데이터의 결과 값  
2) 독립변수: 종속변수에 영향을 주는 데이터 (입력 값)

#### 1) KNN

- KNN(k-Nearest Neighbors)은 지도학습의 한 종류로 종속 변수가 있는 데이터를 사용하여 분류 작업을 하는 알고리즘임. 알고리즘의 이름에서 볼 수 있듯이 데이터로부터 거리가 가까운 k개의 다른 데이터의 정답을 참조하여 분류하는 알고리즘으로 주로 거리를 측정할 때, 유클리디안 거리 계산법<sup>1)</sup>을 사용하여 거리를 측정함
- 예를 들어 새로운 데이터 네모가 있고 이것에 인접한 데이터가 다수 존재한다고 가정할 때 가장 인접한 데이터 5개 중 세모의 비율이 더 높다면 새로운 데이터를 세모로 분류함(그림 5-1 참조)

<그림 5-1> KNN의 분류 방식



출처: <https://superkong1.tistory.com/22>

1) 유클리디안 거리: n차원의 공간에서 두 점간의 거리를 알아내는 공식  
p와 q사이의 거리

$$\|p - q\| = \sqrt{(p - q) \cdot (p - q)} = \sqrt{\|p\|^2 + \|q\|^2 - 2p \cdot q}$$

## 2) 로지스틱 회귀

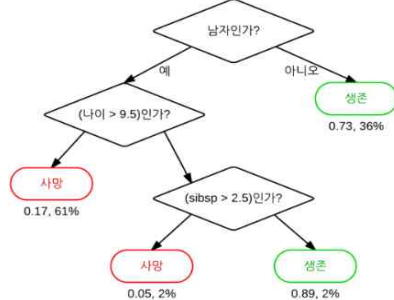
- 로지스틱 회귀는 선형회귀와 마찬가지로 종속변수와 독립변수 간의 관계를 구체적인 함수로 나타내 향후 예측모형에 사용함. 하지만 로지스틱 회귀는 선형회귀와 다르게 종속변수가 범주형 데이터일 때도 유용하며 결과를 특정 카테고리 분류하는데 유용함
- 예를 들어 1과 0을 분류하는 상황일 때 1에 대한 분류만을 0%부터 100%의 확률로 바꾸어 확률이 50% 이상이면 1, 50% 이하이면 0으로 분류함. 그러므로 예측 확률도 선형회귀와 달리 0부터 100까지의 유의한 확률을 나타내고 그 확률을 가지고 50%를 기준으로 분류함

## 3) 의사결정나무

- 의사결정나무는 의사결정규칙을 나무 구조로 나타내어 전체 자료를 몇 개의 소집단으로 분류하거나 예측을 하는 분석방법임
- 상위 노드부터 하위 노드로 트리구조를 형성해 매 단계마다 분류변수와 분류 기준값의 선택이 중요함
- 예를 들어 성별과 나이를 sibsp라는 독립변수로 생존과 사망의 여부 대한 분류를 할 때, 성별이 남자, 나이가 9.5 초과, sibsp의 수치가 2.5이상이면 사망으로 분류하는 것이 불확실성을 가장 감소시키므로 해당 샘플을 사망으로 분류함. 이처럼 의사결정 나무 모형은 각 노드의 특성과 노드가 나타내는 확률에 따라 데이터를 분류·예측하는데 유용함(그림 5-2 참조)

98

<그림 5-2> 의사결정나무 분류 과정 그래프



출처: <https://m.blog.naver.com/2011topcit/220611261399>

## 4) 랜덤 포레스트

- 랜덤 포레스트는 의사결정나무를 기반으로 분류, 회귀분석 등에 자주 사용되는 앙상블 학습 방법<sup>1)</sup>의 일종으로 의사결정나무의 과대적합<sup>2)</sup> 발생 문제점을 개선한 모형임
- 분석에 너무 많은 수의 독립변수가 존재할 경우 과대적합 문제가 발생하기 쉬운데, 랜덤 포레스트는 변수 중 일부만을 취하는 의사결정나무를 다수 생성함, 구성된 다수의 의사결정나무들로부터 각각 결과치들을 도출하여 투표를 실시하며, 가장 많이 표를 얻는 결과를 기반으로 최종 예측값 도출

1) 앙상블 학습 방법: 하나 이상의 머신러닝 모형을 조합하여 분석 정확도를 높이는 분석 기법  
 2) 과대적합: 독립변수가 너무 많을 시, 종속변수에 영향을 크게 끼치지 않고 패턴이 크게 없는 특정 독립변수나 각 독립변수의 패턴이 없는 데이터가 있을 때 그것을 학습할 경우 실질적인 패턴이 약화되어 새로운 데이터를 예측하면 오차율이 증대함



## 5) 에이다 부스팅

- 에이다 부스팅은 간단한 약분류기들을 다수 만들어 상호보완 하도록 단계적으로 학습하며 생성한 모형들을 조합하여 최종적으로 강분류기의 성능을 향상시키는 모형임
- 에이다 부스팅에 사용된 의사결정나무를 기반으로 다수의 약분류기를 생성하여 하나씩 순차적으로 학습을 진행함. 이때 먼저 학습된 분류기가 오분류한 결과정보를 다음 분류기의 학습 시 사용함. 이전 분류기가 오분류한 샘플의 가중치를 주어 오분류된 데이터에 더 집중해 학습하게 함으로써 이렇게 각각의 랜덤 포레스트가 약한 결과를 낸다면 에이다부스팅을 이용한 강분류기는 단일 랜덤 포레스트보다 더 좋은 성능을 보임

## 6) 그라디언트 부스팅

- 그라디언트 부스팅은 모형의 학습과정에서 예측값과 실제값의 차이를 개선해 나가는 모형임
- 의사결정나무를 기반으로 에이다 부스팅과 비슷한 학습 과정을 하지만 그라디언트 부스팅은 오분류 샘플의 가중치를 주는 것이 아닌 모형의 실제 값에 대한 예측 확률이 낮은 샘플을 다음 분류기 학습에 반영해 감으로써 실제값으로 수렴해 가는 모형임

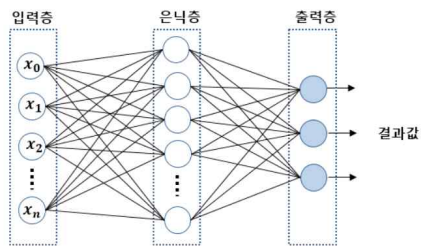
## 7) 다층 퍼셉트론

- 다층 퍼셉트론은 입력층과 출력층 사이에 하나 이상의 중간층이 존재하는 신경망 모형으로 <그림 5-3>에 나타난 것과 같은 계층구조를 가지고 있음
- 입력층과 출력층 사이의 중간층을 은닉층(hidden layer)이라 부름. 신경망은 입력층, 은닉층, 출력층 방향으로 연결되어 있으며 각 층내의 연결과 출력층에서 입력층으로의 직접적인 연결은 존재하지 않는 전방향(Feedforward) 신경망임

1) 약(강)분류기: 설명력이 약한 단순한 형태(설명력이 높고, 복잡한)의 분석 모형

100

<그림 5-3> 다층 퍼셉트론 예측 과정



출처: <https://c11.kr/bm3m>

## 8) 서포트 벡터 머신(SVM)

- 서포트 벡터 머신은 기계학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도학습 모형이며, 주로 분류와 회귀분석을 위해 사용함
- <그림 5-4>와 같이 파란 점의 영역을 1이라 두고 녹색 점의 영역을 -1이라 두고 이것을 초평면<sup>1)</sup>으로 분류함. 이때 데이터를 가장 잘 분류해주는 초평면을 찾기 위해 초평면에서 가장 가까운 양측의 데이터 포인트 즉, 서포트 벡터들과 초평면의 거리(margin)을 가장 크게 하는 지점을 찾게 되고, 그 선을 기준으로 분류 등의 분석을 하는 모형임
- Margin을 가까운 데이터 포인트로부터의 거리라고 정의하고, 이를  $x^+$ ,  $x^-$ 로 각각 정의함

$$Margin = distance(x^+, x^-) = \|x^+ - x^-\|_2 \quad \text{<수식 5-1>}$$

1) 초평면: 어떤 공간이 있을 때, 이 공간의 한 점을 통과하는 해집합을 말함

- 이때, 양평면의 점  $x^+$ 가 음평면의 점  $x^-$ 으로 평행이동한 값을 수식으로 나타내면

$$x^+ = x^- + \lambda w \text{ 이고,}$$

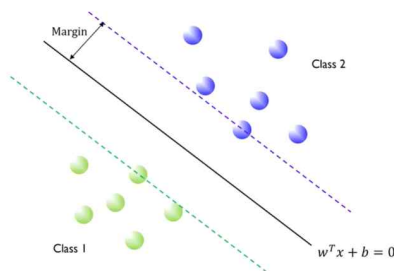
- 이를 <수식 5-1>에 대입하면

$$\begin{aligned} &= \| (x^- + \lambda w) - x^- \|_2 \\ &= \| \lambda w \|_2 \\ &= \lambda \sqrt{w^T w} = \frac{2}{w^T w} \cdot \sqrt{w^T w} = \frac{2}{\sqrt{w^T w}} = \frac{2}{\|w\|_2} \text{ 이 도출됨} \end{aligned}$$

- 이에 Margin이 크면 클수록 인접 데이터와의 거리가 멀어진다는 점을 가정하여 Margin을 최대화 해주는 w값을 찾음으로써 분류 등의 분석을 진행함 <수식 5-3>

$$\max \text{Margin} = \max \frac{2}{\|w\|_2} = \min \frac{1}{2} \|w\|_2 \text{ <수식 5-3>}$$

<그림 5-4> SVM 분류 방식



출처: <https://c11.kr/bm3q>

1) w는 초평면의 기울기 계수를 의미하며, 즉 변수의 가중치라고 생각할 수 있음

## 다. 업체추천 분석 활용변수

### 1) 업체특성

- 조사지역 내 업체들의 특성을 반영하기 위해 농관원 내부자료를 활용하여 업종, 재적발달을 도출하며 도출된 변수는  $100m^2$  그리드 단위로 계산함
- 업체의 업종은 음식점업, 소매업, 유통, 식료품가공업, 도매업, 급식업, 그리고 생산업을 기준으로 분류함
- 재적발달은 재적발 여부를 기준으로 분류함
- 머신러닝 모형에 사용된 데이터는 기존 전국 업체 단위 데이터를 사용함. 기존 전국 업체 단위 데이터의 주소 정제 과정 중 누락된 데이터가 있어 일부를 제외하였음. 따라서 기존 전국 업체 단위 데이터 단속률과 정제한 데이터 단속률 간에 약간의 차이가 발생됨

### 2) 지역특성

- 조사지역 자체의 특성을 분석에 반영하기 위해 공공데이터를 활용함
- 평균 폐업일은 동 단위자료를 취합하여 해당 그리드에 반영함
- 평균 거리 변수의 경우  $100m^2$  그리드 내 조사업체와 다른 지역 업체(축산판매업, 대학, 유통전문업, 위탁 급식, 초등학교, 중학교, 고등학교, 병원, 전국 시장, 유흥업소, 경유점, 유치원, 어린이 보호구역, 식품 가공업, 음식점)간의 거리 평균을 계산함

<표 5-1> 활용변수표

대분류	중분류	내용
업체특성	제적발 여부	제적발 여부
	업종	음식점업, 소매업, 유통업, 식료품 가공업, 도매업, 급식업, 생산업
지역특성	평균폐업일	행정동별 자료
	평균 거리	반경 500m내(축산판매업, 대학, 유 통전문업, 위탁 급식, 초등학교, 중 학교, 고등학교, 병원, 전국 시장, 유흥업소, 정육점, 유치원, 어린이 보호구역, 식품 가공업, 음식점)

## 라. 적용과정

### 1) 전처리

- 먼저 데이터를 모형에 학습시킬 훈련데이터와 그 결과를 확인해 볼  
검정데이터로 나누어서 진행함
- 2009년부터 2017년까지를 훈련데이터로 2018년을 검정데이터로 함

<그림 5-5> 전처리하기 전 데이터

작업예부 ID	Dist	축산판매	도매	소매	정육점	유통	음식점	NEAR	DIST	거리	DIST	식품	DIST	병원	DIST	대학	NEAR	DIST	유통	DIST	음식점	DIST	전국	제적발여부	제적발일	평균폐업일
1	1	19	0.005674	0.003417	0.001198	0.000505	0.000505	0.000975	0.000975	0.08548	0.001394	0.027477	0.027477	0.016509	0.003442	0.022681	0.028147	0	1	2157						
3	2	31	0.001837	0.003381	0.000338	0.005942	0.005942	0.00319	0.00319	0.13243	0.003241	0.015181	0.015181	0.004504	0.004994	0.00239	0.009163	0.001963	0	1	2157					
4	3	67	0.000386	0.001588	0.000038	0.000878	0.000878	0.002915	0.002915	0.078239	0.006553	0.022119	0.022119	0.002686	0.005085	0.004654	0.020012	0.001202	0	1	2157					
5	4	68	0.002492	0.002015	0.001709	0.005608	0.005608	0.003604	0.003604	0.204915	0.004946	0.052194	0.052194	0.006769	0.011949	0.011851	0.038189	0.0019	0	1	2157					
6	5	76	0.004624	0.00612	0.004624	0.003453	0.003453	0.005177	0.005177	0.131703	0.024027	0.063066	0.063066	0.001817	0.003457	0.00621	0.031287	0.001202	0	1	2008					
7	6	85	0.003882	0.00331	0.004238	0.00397	0.00397	0.00597	0.00597	0.001472	0.001472	0.077449	0.006482	0.013845	0.013845	0.000279	0.007813	0.001341	0	1	2008					
8	7	110	0.000712	0.001269	0.000417	0.001617	0.001617	0.00238	0.00238	0.115865	0.001172	0.002443	0.002443	0.001524	0.002397	0.001538	0.006077	0.0007	0	1	2008					
9	8	160	0.007788	0.0032	0.00481	0.005702	0.005702	0.003643	0.003643	0.001699	0.016894	0.021074	0.021074	0.002947	0.016695	0.003232	0.027938	0.0007	0	1	2008					
10	9	196	0.003735	0.001172	0.001239	0.001744	0.001744	0.001359	0.001359	0.059088	0.001781	0.013736	0.013736	0.003754	0.005796	0.013736	0.002717	0.0007	0	1	2047					
11	10	197	0.004112	0.00553	0.003399	0.003834	0.003834	0.001359	0.001359	0.059088	0.001781	0.013736	0.013736	0.003754	0.005796	0.013736	0.002717	0.0007	0	1	2008					
12	11	198	0.002415	0.001817	0.001329	0.00311	0.00311	0.001231	0.001231	0.176552	0.000852	0.030149	0.030149	0.003063	0.003028	0.016829	0.037219	0.0007	0	1	2008					
13	12	212	0.00333	0.002595	0.001788	0.007944	0.007944	0.002037	0.002037	0.098822	0.001564	0.03026	0.03026	0.009617	0.009162	0.014554	0.013259	0.0007	0	1	2047					
14	13	215	0.00784	0.001072	0.003697	0.009372	0.009372	0.002655	0.002655	0.002472	0.05796	0.05796	0.003002	0.006342	0.006297	0.001989	0.0007	0	1	2008						
15	14	216	0.001664	0.005947	0.001026	0.000318	0.000318	0.003521	0.003521	0.136188	0.003121	0.014983	0.014983	0.002817	0.007513	0.000781	0.000927	0.0007	0	1	2157					
16	15	217	0.007669	0.00466	0.000383	0.000469	0.000469	0.000888	0.000888	0.018449	0.003993	0.028574	0.028574	0.003749	0.006458	0.018242	0.017484	0.0007	0	1	2157					
17	16	218	0.000562	0.00174	0.000896	0.005527	0.005527	0.004144	0.004144	0.023879	0.005288	0.023891	0.023891	0.006292	0.019379	0.00171	0.0007	0	1	2261						
18	17	219	0.002795	0.000533	0.001375	0.001816	0.001816	0.003565	0.003565	0.12406	0.000752	0.011623	0.011623	0.004625	0.00919	0.018244	0.006919	0.0007	0	1	2711					

- <그림 5-5>를 보게 되면 데이터 각 변수들의 수치가 큰 폭으로 차이가 나므로 이 데이터들을 분석에 용이하도록 정규화를 진행함

- MinMaxScaler 함수를 사용하여 데이터 셋 안의 데이터 값 중 가장 큰 값을 1, 가장 적은 값을 0으로 기준을 정해서 데이터의 각 값들을 0과 1사이의 값으로 변환하여 분석이 가능한 형태로 정규화함

## 2) 모형 적합 및 훈련

- 전처리가 끝난 데이터를 KNN, 로지스틱 회귀, 의사결정나무, 그래디언트 부스팅, 에이다 부스팅, 랜덤 포레스트, 다층 퍼셉트론, SVM에 적합 시킴
  - KNN에서는 K를 8개로 잡아서 훈련함
  - 로지스틱 회귀와 의사결정나무와 SVM은 파라미터 값을 기본값으로 설정하여 훈련함
  - 그래디언트 부스팅, 에이다 부스팅, 그리고 랜덤 포레스트는  $n\_estimators=10$ 으로 설정하여 훈련함
  - 다층 퍼셉트론은 은닉층을 1층으로 설정하여 훈련함

1)  $n\_estimators$ : 부스팅 모형에서는 결합할 분류기의 개수 지정하는 파라미터

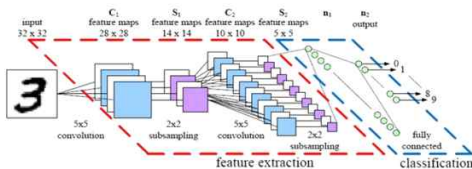
106

## 2 딥러닝(CNN)을 활용한 단속 지역추천

가. 모형 개요<sup>1)</sup>

- CNN(Convolutional Neural Network)는 머신러닝의 한 유형인 딥러닝(Deep Learning)<sup>2)</sup>으로 스스로 이미지, 영상, 음향의 특성을 학습하여 분류하는데 탁월한 성능을 보이며 가장 많이 사용되는 알고리즘 중 하나임. 특히 CNN은 이미지나 음성 텍스트에서 수치적 배열의 패턴을 찾는데 특히 유용함
- CNN은 네트워크를 구성하는 수십, 수백 개의 계층이 각각 이미지의 서로 다른 특징을 찾으려 학습할 수 있음
  - 필터가 각 학습 이미지에 서로 다른 형태로 적용되고 필터의 출력은 다음 계층의 입력으로 활용됨. 필터는 뚜렷한 부분 및 가장자리 등과 같이 매우 단순한 특징에서 시작하여 객체만의 고유한 특징으로 더 복잡하게 발전할 수 있음. 따라서 CNN은 각 데이터의 특징을 식별하고 분류하는데 탁월한 성능을 보임

<그림 5-6> 3이 적혀있는 이미지를 CNN 계층을 통해 분류



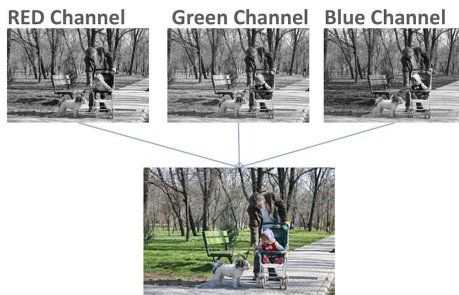
출처: <https://eehokrap.tistory.com/281>

1) "CNN, Convolutional Neural Network 요약" TAEWAN.KIM 블로그, 2018년 1월 4일 수정, 2019년 11월 18일 접속, <http://taewan.kim/post/cnn/>.  
 2) 딥러닝(Deep Learning): 딥러닝 또는 심층학습은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화를 시도하는 기계학습 알고리즘의 집합으로 정의되며, 큰 틀에서 삶의 사고방식을 컴퓨터에게 가르치는 기계학습의 한 분야라고 이야기할 수 있음 출처(<https://bit.ly/2XsL015>)

## 나. CNN 구성

- 합성곱 계층(Convolutional Layer): Convolutional Layer는 입력 데이터로부터 특징을 추출하는 역할을 수행
  - 이는 특징을 추출하는 필터(Filter)와 필터의 값을 비선형 값으로 바꾸어주는 활성화 함수(Activation Function)로 이루어져 있음
- 채널(Channel): 이미지 픽셀 하나하나를 실수이며 컬러사진은 색을 표현하기 위해 각 픽셀을 RGB 3개의 실수로 표현한 3차원 데이터이고 흑백의 이미지인 경우 1개의 채널로 구성됨. Convolutional Layer에 입력되는 데이터는 한 개 이상의 필터가 적용됨
  - 1개의 필터는 Feature Map의 채널이 되며 Convolutional Layer에 n개의 필터에 적용된다면 출력 데이터는 n개인 채널을 갖게 됨

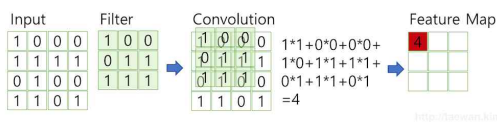
<그림 5-7> 33개의 채널로 만들어진 컬러사진



출처: <http://taewan.kim/post/cnn>

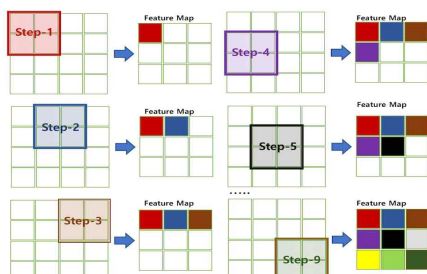
- 필터(Filter): 스트라이드(Stride) 필터는 이미지의 특징을 찾아내기 위한 파라미터이며 필터는 일반적으로 정사각형 행렬로 정의됨
  - CNN에서 학습의 대상은 필터 파라미터임. 필터는 입력 데이터를 지정한 간격으로 이동하면서 합성곱 계산을 하며 여기서 지정된 간격을 Stride라고 함. Stride가 1이면 1칸 만큼 이동하고 2이면 2칸 만큼 이동을 함. Convolution Layer의 입력 데이터를 필터가 이동하면서 합성곱을 통해 만든 출력을 Feature Map이라고 하며 이 Feature Map 행렬에 활성화 함수를 적용한 결과가 Activation Map임

<그림 5-8> 2차원 입력 데이터를 1개의 필터로 합성곱 계산을 수행



출처: <http://taewan.kim/post/cnn>

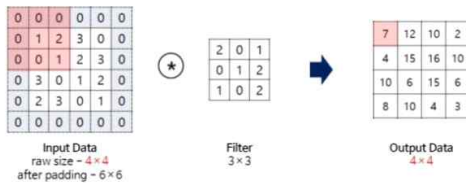
<그림 5-9> Stride가 1로 필터를 입력 데이터에 이동하는 과정



출처: <http://taewan.kim/post/cnn>

- 패딩(Padding): Convolution Layer에서 Filter와 Stride에 작용하여 Feature Map 크기는 입력 데이터보다 작음. 이러한 Convolution Layer 계산에서 데이터의 크기가 줄어들어 데이터가 손실되는 것을 방지하는 것이 패딩임
- 패딩은 입력 데이터 외곽지역에 지정된 픽셀만큼 특정 값으로 채워 넣는 것을 의미하며 보통 패딩값으로 0을 사용함

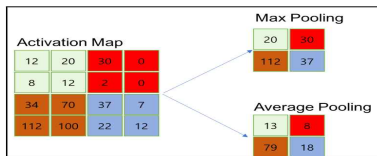
<그림 5-10> 데이터의 외곽에 패딩을 사용하여 데이터 크기를 유지



출처: <https://eehoeskraptistory.com/281>

- 풀링(Pooling): 풀링은 Convolutional Layer의 출력 데이터를 입력 받아서 출력 데이터(Activation Map)의 크기를 줄이거나 특정 데이터를 강조하는 용도로 사용함
- 풀링을 처리하는 방법으로는 MaxPooling과 AveragePooling이 있으며 정사각형 행렬의 특정 영역 안에 값의 최댓값을 찾거나 특정 영역의 평균값을 구하는 방식으로 작동함

<그림 5-11> MaxPooling과 AveragePooling의 예시



출처: <http://taewan.kim/post/cnn>

## 다. 지역추천 분석 활용변수

### 1) 업체특성

- 조사지역 내 업체들의 특성을 반영하기 위해 농관원 내부 자료를 활용하여 취급품목, 업종, 연령별 업체수를 도출하였으며 도출된 변수는  $100m^2$  그리드 단위로 계산함
- 취급 품목별 업체 수는 배추김치, 쌀, 돼지고기, 닭고기 그리고 쇠고기 5개 품목을 기준으로 계산한 업체수를 사용함
- 업종별 업체수는 음식점업, 소매업, 유통업, 식료품가공업, 도매업, 급식업, 그리고 생산업 기준으로 계산함
- 연령별 업체수는 각 20대, 30대, 40대, 50대 그리고 70대 이상으로 분류하여 계산함

### 2) 지역특성

- 조사지역 자체의 특성을 분석에 반영하기 위해 공공데이터를 활용함
- 공무원수, 소득세<sup>1)</sup>, 면적, 인구, 인구밀도의 경우 행정동 단위자료를 취합하여 해당 그리드에 반영함
- 평균 거리 변수의 경우  $100m^2$  그리드 내 조사업체와 다른 지역업체(대학, 우수업체, 유통센터, 급식, 학교, 병원, 축산유통, 유통, 시장, 유흥업소)간의 거리 평균을 계산함

### 3) 외부특성

- 국산가격, 수입가격, 수입물량(돼지, 쇠고기, 배추김치, 쌀, 닭고기), 돼지고기, 쇠고기, 배추김치, 쌀, 닭고기 관련 뉴스 스코어<sup>2)</sup>
- 외부특성은 월별 해당 품목의 가격변화와 물량변화, 뉴스 스코어를 수치로 사용

1) 소득세의 경우 가용데이터 중 최신 연도인 2012년 자료 활용  
2) 뉴스 스코어: 월별 품목에 관한 뉴스의 빈도를 점수로 변환

<표 5-2> 활용 변수표

대분류	중분류	내용
업체특성	품목별 취급 업체수	배추김치, 쌀, 닭고기, 돼지고기, 쇠고기
	업종별 업체수	음식점업, 소매업, 유통업, 식료품가공업, 도매업, 급식업, 생산업
	연령별 업체수	20대, 30대, 40대, 50대, 60대, 70대 이상
	성별 업체수	남, 여
지역특성	공무원수	행정동별 자료
	소득세	
	면적	
	인구	
	인구밀도	
	평균 거리	반경 500m내 (대학, 급식, 유치원, 학교, 병원, 축산유통, 유통, 시장, 유흥업소)
외부요소	지역 내 업체수	위반 업체수
	국산가격	1~12월, 월별 배추김치, 쌀, 돼지고기, 쇠고기, 닭고기
	수입가격	
	수입물량	
	뉴스 스코어	

## 라. CNN 모형 적용과정

딥러닝(CNN) 모형을 통한 전처리 부분은 머신러닝을 활용한 단속업체 적용과정과 거의 유사하나 사용되는 변수가 다름. 단속업체 적용과정은 머신러닝 모형에 적합하기 위해 그에 맞는 형태로 바꾸는 것이고, 단속 지역 적용과정은 CNN 모형에 맞게 형태를 바꾸는 것임.

### 1) 전처리

- 데이터를 모형에 학습시킬 훈련데이터와 그 결과를 확인하기 위한 검정데이터로 나눔
- 2009년부터 2017년까지를 훈련데이터로 정하고 2018년을 검정데이터로 정해서 분석을 진행함

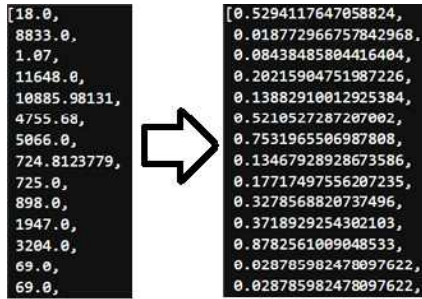
<그림 5-12> 전처리 전 데이터

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연	MEAN, 연
2	0	0	0	8205.76	4342	4091.56	4092	2739	4461	2199	2397	2397	1	63.333333	0
3	1.07	11648	10885.98	4755.68	5066	724.8124	725	898	1947	3204	69	69	0	0	0
4	1.07	11648	10885.98	4695.356	5092.588	769.412	769.0588	890.3529	1977.529	3224.882	21.70588	21.70588	1	15.88235	0
5	1.07	11648	10885.98	4695.356	5092.588	769.412	769.0588	890.3529	1977.529	3224.882	21.70588	21.70588	1	15.88235	0
6	1.07	11648	10885.98	4695.356	5092.588	769.412	769.0588	890.3529	1977.529	3224.882	21.70588	21.70588	1	15.88235	0
7	1.07	11648	10885.98	4695.356	5092.588	769.412	769.0588	890.3529	1977.529	3224.882	21.70588	21.70588	1	15.88235	0
8	1.07	11648	10885.98	4662.34	5110	798.2968	798	891	1998	3239	54	54	0	0	0
9	12.68	18549	1462.855	7556.82	3690	3438.09	3438	2103	4274	1548	1746	1746	0	0	0
10	12.68	18549	1462.855	7772.91	3901	3640.85	3641	2244	4116	1771	1964	1964	0	0	0
11	1.07	11648	10885.98	4774.06	4845	489.4646	489	766	1721	2993	296	296	3	29	0
12	1.07	11648	10885.98	4627.08	4687	375.381	375	576	1570	2825	413	413	0	55	0
13	1.07	11648	10885.98	4729.36	4510	152.1196	152	596	1385	2665	341	341	1	53	0
14	1.07	11648	10885.98	4569.038	4470.222	223.5483	223.5556	439	1354.556	2608.222	292.6667	292.6667	0	15.444444	0
15	1.07	11648	10885.98	4569.038	4470.222	223.5483	223.5556	439	1354.556	2608.222	292.6667	292.6667	0	15.444444	0
16	1.07	11648	10885.98	4569.038	4470.222	223.5483	223.5556	439	1354.556	2608.222	292.6667	292.6667	0	15.444444	0
17	1.07	11648	10885.98	4483.52	4438	283.4799	283	352	1332	2568	256	256	13	25.74074	0
18	1.71	37144	21721.64	4802.97	4410	73.69978	74	643	1287	2579	298	298	0	0	0
19	1.07	11648	10885.98	4729.22	4354	11.37653	11	580	1229	2515	215	215	0	0	0

- <그림 5-12>을 보면 전처리가 되기 전 상태는 각 변수들의 수치가 큰 폭으로 차이가 나므로 이 데이터들을 분석에 용이하도록 정규화 하여야 하며 이를 위해 정규화 방법 중 하나인 MinMaxScaler<sup>1)</sup>를 사용함. MinMaxScaler를 사용하여 데이터를 변환 후 CNN 분석을 위해 2차원 데이터 프레임을 3차원 형태로 한번 더 변환함

1) MinMaxScaler: 최대/최소값이 각각 1, 0이 되도록 스케일링

<그림 5-13> MinMaxScaler사용 전과 후

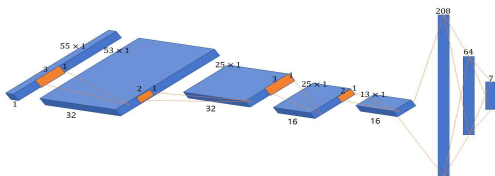


- $100m^2$  그리드 단위를 1개 지역으로 각 지역의 위험등급은 지난 10년간 그 지역에서 발생한 단속여부의 합계를 기준으로 정하여 이것을 종속변수로 사용함
- 해당 지역의 누적합계를 기준으로 데이터 정제를 위해 Natural Break 기법을 활용하여 0-6 사이의 군집으로 정제하였으며 이를 위험도로 정의함

## 2) 모형 적합 및 훈련

- 전처리가 다 끝난 데이터를 CNN에 적합 시킴. 본 추천서비스에 사용될 CNN의 형태는 Convolutional Layer를 2층으로 하는 모형임

<그림 5-14> CNN 2층 구상도



- Convolution Layer 1층: 데이터는  $55 \times 1$ 의 행렬로, 이를  $3 \times 1$  크기인 필터로 합성곱을 계산함. 이를 통해 32개의 Feature Map을 생성하며 Convolution Layer를 거친 데이터는 다음 식을 통해 크기가 결정됨

$$OutputHeight = \frac{H + 2P - FH}{S} + 1$$

$$OutputWeight = \frac{W + 2P - FW}{S} + 1$$

- OutputHeight는 출력 데이터 높이를 의미하며 OutputWeight는 출력 데이터 폭을, H는 입력 데이터 높이를, 그리고 W는 입력 데이터 폭을, FH는 필터 높이를, 그리고 FW는 필터 폭을, S는 Strid 크기를, P는 패딩 사이즈를 의미함

- 풀링(Pooling) 1층: 마찬가지로  $53 \times 1$  데이터를  $2 \times 1$  크기 풀링사이즈로 MaxPooling 계산함. 이 과정에서 데이터는 원래 크기의  $\frac{1}{2}$ 로 감소

- Convolution Layer 2층에서는 25개의 데이터를  $3 \times 1$  크기의 필터가 이동하면서 합성곱 계산을 진행하며 결과적으로 16개의 Feature Map을 생성함



- 이후 풀링(Pooling)과정을 거쳐 합성곱 계산이 마무리되고 결과 데이터가 Flatten Layer<sup>1)</sup>단계로 넘어감. 총 3개의 Flatten Layer를 거쳐 0-6 사이의 위험도 예측값을 부여함
- 이렇게 55개의 변수를 가진 데이터로 CNN을 사용한 분석이 끝남

1) Flatten Layer: 다차원 데이터를 1차원으로 변환시켜주는 층

## VI 분석결과

### 1 단속 업체추천

가. 머신러닝을 활용한 업체추천 결과

#### 1) 업체추천 결과(서울특별시)

- 잠재적 위반의 가능성의 업체를 추천해주는 데이터의 종속변수인 적발여부는 실제 단속원들이 조사를 나갔던 곳의 실제 결과임
- 모형의 예측값과 그것의 실제값과의 정확도 성능평가는 f1스코어<sup>1)</sup>를 가지고 위반에 대한 예측 정확도로 함

<표 6-1> 서울특별시 업체추천 모형별 f1스코어

모형	f1스코어
KNN	0.071
로지스틱 회귀	0.159
의사결정나무	0.054
그래디언트 부스팅	0.16
에이다 부스팅	0.16
랜덤 포레스트	0.07
다층 퍼셉트론	0.107
서포트 벡터 머신	0.16

1) f1스코어: 정밀도와 재현율의 기중조화평균(weight harmonic average)을 f점수(f-score)라고 함. 정밀도(precision)는 양성 클래스(1-위반)에 속한다고 출력한 샘플 중 실제로 양성 클래스에 속하는 샘플 수의 비율을 말하며 재현율(recall)은 실제 양성 클래스에 속한 표본 중에 양성 클래스에 속한다고 출력한 표본의 수의 비율을 뜻함

- <표 6-1>은 서울특별시의 2018년 데이터에 대해 각 모형들을 가동하여 f1스코어를 산출함
- <표 6-1>에서 높은 성능을 보인 모형들은 그래디언트 부스팅, 에이다 부스팅, 서포트 벡터 머신임
- 모형 3개가 같은 점수를 보였는데, 그중 서포트 벡터 머신을 사용함
  - 이유는 부스팅 모형들은 처음 훈련데이터를 학습하고 훈련데이터를 예측했을 때, 발생하는 잘못된 예측들의 오차가 있는데 그 오차 만큼에 가중치를 두어서 다음번에 학습하여 훈련 오차를 줄임. 이것은 훈련데이터의 예측 오차는 줄일 수 있지만 과대적합으로 검증데이터의 예측 오차가 커질 수 있음. 하지만 서포트 벡터 머신은 오차에 대해서 가중치를 두고 다시 학습하는 방법이 아님. 따라서 부스팅의 과대적합의 문제를 피할 수 있어 서포트 벡터 머신을 모형으로 선정함

<표 6-2> 분류결과표(Confusion\_Matrix)

실제 \ 모형	미위반(0)	위반(1)	총합
미위반(0)	17,334	-	17,334
위반(1)	378	36	414
총합	17,712	36	-

- <표 6-2>를 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시됨. 모형이 미위반이라고 예측한 개수는 총 17,712개 임. 그 중에서 실제 위반 개수 17,334개를 모두 맞추었고 틀린 것은 378개임. 그리고 모형이 위반이라고 예측한 것은 총 36개로 예측함. 실제 위반데이터 414개 중에서 36개만 예측하고 맞춤

<표 6-3> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.98	1.00	0.99	17,334
위반(1)	1.00	0.09	0.16	414
micro avg	0.98	0.98	0.98	17,748
macro avg	0.99	0.54	0.57	17,748
weighted avg	0.98	0.98	0.97	17,748

- <표 6-3>는 위반과 미위반의 예측 정확도를 정리한 것임. 미위반에 대한 f1스코어는 0.99임, 이에 대한 설명으로 모형은 미위반에 대한 예측을 99% 정도의 정확성으로 예측하여 미위반 업체를 피해서 갈 수 있기에 조사의 효율성을 제고시킴
- 위반에 대한 f1스코어는 0.16임, 즉 서울특별시의 단속원들이 조사를 나갈 시, 9년 동안의 경험을 근거한 2018년 단속률은 2.3% 밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률은 16%까지 향상될 것으로 봄

## 2) 업체추천 결과(대전광역시)

<표 6-4> 대전광역시 업체추천 모형별 f1스코어

모형	f1-score
KNN	0.089
로지스틱 회귀	0.125
의사결정나무	0.062
그래디언트 부스팅	0.125
에이다 부스팅	0.125
랜덤 포레스트	0.106
다층 퍼셉트론	0.127
서포트 벡터 머신	0.125

- <표 6-4>에서는 대전광역시의 2018년 데이터를 각 모형으로 가동하였음. 가동한 모형 중 f1스코어가 가장 높게 나와 우수한 성능을 보인 모형은 다층 퍼셉트론이며 따라서 다층 퍼셉트론 기법을 선정함

<표 6-5> 분류결과표(Confusion\_Matrix)

실제 \ 모형	미위반(0)	위반(1)	총합
미위반(0)	7,077	55	7,132
위반(1)	177	17	194
총합	7,254	72	-

- <표 6-5>을 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시됨. 모형이 미위반이라고 예측한 것은 총 7,254게임. 이 중에서 맞춘 개수는 7,077개이고 틀린 것은 177게임

120

그리고 모형이 위반이라고 예측한 것은 총 72게임. 이 중에서 맞춘 것은 17개이고 틀린 것은 55게임

<표 6-6> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.98	0.99	0.98	7,132
위반(1)	0.24	0.09	0.13	194
micro avg	0.97	0.97	0.97	7,326
macro avg	0.61	0.54	0.56	7,326
weighted avg	0.96	0.97	0.96	7,326

- <표 6-6>는 위반과 미위반의 예측 정확도를 정리한 것으로 미위반에 대한 f1스코어는 0.98임. 이에 대한 설명으로 모형은 미위반에 대한 예측을 98% 정도의 정확성으로 예측함
- 위반에 대한 f1스코어는 반올림하여 0.13임. 즉, 대전광역시의 단속 원들이 조사를 나갈 시 9년 동안의 경험을 근거한 2018년 단속률은 2.6%밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률은 13%까지 향상될 것으로 봄

121

3) 업체추천 결과(광주광역시)

<표 6-7> 광주광역시 업체추천 모형별 f1스코어

모형	f1-score
KNN	0.058
로지스틱 회귀	0.057
의사결정나무	0.056
그래디언트 부스팅	0.067
에이다 부스팅	0.116
랜덤 포레스트	0.069
다중 퍼셉트론	0.071
서포트 벡터 머신	0.0608

- <표 6-7>에서는 광주광역시의 2018년 데이터를 각 모형으로 가동하였음. 가동한 모형 중 f1스코어가 가장 높게 나와 우수한 성능을 보인 모형은 에이다 부스팅이며 에이다 부스팅 기법을 선정함

<표 6-8> 분류결과표(Confusion\_Matrix)

실제 \ 모형	미위반(0)	위반(1)	총합
미위반(0)	6,350	330	6,680
위반(1)	139	31	170
총합	6,489	361	-

- <표 6-8>을 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시됨. 모형이 미위반이라고 예측한 개수는 총 6,489개이며 이 중에서 맞춘 것은 6,350개이고 틀린 개수는 139개임.

모형이 위반에 대해서 예측한 것은 총 361개를 예측하였으며 이 중에서 31개를 맞추고 330개를 틀림

<표 6-9> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.98	0.95	0.96	6,680
위반(1)	0.09	0.18	0.12	170
micro avg	0.93	0.93	0.93	6,850
macro avg	0.53	0.57	0.54	6,850
weighted avg	0.96	0.93	0.94	6,850

- <표 6-9>는 위반과 미위반의 예측 정확도를 정리한 것으로 미위반에 대한 f1스코어는 0.96임. 이에 대한 설명으로 미위반에 대한 예측을 96%의 정확성으로 예측함
- 위반에 대한 f1스코어는 반올림하여 0.12임. 즉, 광주광역시의 단속원들이 조사를 나갈 시 9년 동안의 경험을 근거한 2018년 단속률은 2.4% 밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률은 12%까지 향상될 것으로 봄

#### 4) 업체추천 결과(인천광역시)

<표 6-10> 인천광역시 업체추천 모형별 f1스코어

모형	f1-score
KNN	0.145
로지스틱 회귀	0.318
의사결정나무	0.088
그래디언트 부스팅	0.323
에이다 부스팅	0.323
랜덤 포레스트	0.137
다중 퍼셉트론	0.157
서포트 벡터 머신	0.323

- <표 6-10>에서는 인천광역시의 2018년 데이터를 각 모형으로 가동하였음. 가동한 모형 중 f1스코어가 가장 높은 성능을 보인 모형은 그래디언트 부스팅, 에이다 부스팅, 서포트 벡터 머신임

- 그중 서포트 벡터 머신을 사용함

- 이유는 부스팅 모형들은 처음 훈련데이터를 학습하고 훈련데이터를 예측했을 때 발생하는 잘못된 예측들의 오차가 있는데 그 오차 만큼에 가중치를 두어서 다음번에 학습하여 훈련 오차를 줄임. 이것은 훈련데이터의 예측 오차는 줄일 수 있지만 과대적합으로 검증데이터의 예측 오차가 커질 수 있음. 하지만 서포트 벡터 머신은 오차에 대해서 가중치를 두고 다시 학습하는 방법이 아님. 따라서 부스팅의 과대적합의 문제를 피할 수 있어 서포트 벡터 머신을 모형으로 선택함

124

<표 6-11> 분류결과표(Confusion\_Matrix)

모형 실제	미위반(0)	위반(1)	총합
미위반(0)	4,889	-	4,889
위반(1)	88	21	109
총합	4,977	21	-

- <표 6-11>을 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시됨. 모형이 미위반이라고 예측한 개수는 총 4,977개이며 이 중에서 맞춘 것은 4,889개이고 틀린 개수는 88개임. 위반에 대해서는 21개만 예측함. 전체 위반데이터 개수 109개 중에서 21개만 예측하고 맞춤

<표 6-12> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.98	1.00	0.99	4,889
위반(1)	1.00	0.19	0.32	109
micro avg	0.98	0.98	0.98	4,998
macro avg	0.99	0.60	0.66	4,998
weighted avg	0.98	0.98	0.98	4,998

- <표 6-12>은 위반과 미위반의 예측 정확도를 정리한 것이며 미위반에 대한 f1스코어는 0.99임

- 이에 대한 설명으로 모형은 미위반에 대한 예측을 99% 정확성으로 예측하여 미위반 업체를 피해서 갈 수 있기에 조사의 효율성을 제고 시킴

- 위반에 대한 f1스코어는 0.32임. 즉, 인천광역시의 단속원들이 조사를 나갈 시 9년 동안의 경험을 근거한 2018년 단속률은 2.2% 밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률을 32%까지 향상될 것으로 봄

125

## 5) 업체추천 결과(부산광역시)

<표 6-13> 부산광역시 업체추천 모형별 f1스코어

모형	f1-score
KNN	0.045
로지스틱 회귀	0.056
의사결정나무	0.050
그래디언트 부스팅	0.131
에이다 부스팅	0.131
랜덤 포레스트	0.082
다중 퍼셉트론	0.058
서포트 벡터 머신	0.057

- <표 6-13>에서는 부산광역시의 2018년 데이터를 각 모형으로 가동하였음. 가동한 모형 중 f1스코어가 가장 우수한 성능을 보인 모형은 에이다 부스팅과 그래디언트 부스팅임
- 2개의 모형이 좋은 성능이 나옴. 그중 그래디언트 부스팅을 사용함
  - 이유는 에이다 부스팅은 잘못 분류된 데이터에 가중치를 두고 학습함. 만일 그 데이터가 패턴이 전혀 없는 경우 모형은 패턴이 없는 것도 가중치를 두고 학습을 하기에 성능이 더욱 약해짐. 따라서 그래디언트 부스팅을 모형으로 선정함

126

<표 6-14> 분류결과표(Confusion\_Matrix)

모형	미위반(0)	위반(1)	총합
실제			
미위반(0)	10,612	-	10,612
위반(1)	211	16	227
총합	10,823	16	

- <표 6-14>을 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시되며 모형이 미위반이라고 예측한 개수는 총 10,823개임. 이 중에서 맞춘 것은 10,612개이고 틀린 개수는 211개임. 위반에 대해서는 16개만 예측함. 전체 실제 위반데이터 개수 227개 중 16개만 예측하고 맞춤

<표 6-15> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.98	1.00	0.99	10,612
위반(1)	1.00	0.07	0.13	227
micro avg	0.98	0.98	0.98	10,839
macro avg	0.99	0.54	0.56	10,839
weighted avg	0.98	0.98	0.97	10,839

- <표 6-15>은 위반과 미위반의 예측 정확도를 정리한 것이며 미위반에 대한 f1스코어는 0.99임
- 이에 대한 설명으로 모형은 미위반에 대한 예측을 99% 정확성으로 예측함
- 위반에 대한 f1스코어는 0.13임. 즉, 부산광역시의 단속원들이 조사를 나갈 시 9년 동안의 경험을 근거한 2018년 단속률은 2% 밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률은 13%까지 향상될 것으로 봄

127

6) 업체추천 결과(대구광역시)

<표 6-16> 대구광역시 업체추천 모형별 f1스코어

모형	f1-score
KNN	0.099
로지스틱회귀	0.194
의사결정나무	0.083
그래디언트 부스팅	0.184
에이다 부스팅	0.182
랜덤 포레스트	0.119
다층 퍼셉트론	0.133
서포트 벡터 머신	0.204

○ <표 6-16>에서는 대구광역시의 2018년 데이터를 각 모형으로 가동하였음. 가동한 모형 중 f1스코어가 가장 우수한 성능을 보인 모형은 서포트 벡터 머신이며 따라서 서포트 벡터 머신 기법을 선정함

<표 6-17> 분류결과표(Confusion\_Matrix)

모형 실제	미위반(0)	위반(1)	총합
미위반(0)	4,291	43	4,334
위반(1)	152	25	177
총합	4,443	68	

○ <표 6-17>을 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시되며 모형이 미위반이라고 예측한 개수

128

는 총 4,443게임. 이 중에서 맞춘 것은 4,291개이고 틀린 개수는 152개임. 위반에 대해서는 68개만 예측함. 전체 실제 위반데이터 개수 177개 중 맞춘 것은 25개이고 틀린 개수는 43개임

<표 6-18> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.97	0.99	0.98	4,334
위반(1)	0.37	0.14	0.20	177
micro avg	0.96	0.96	0.96	4,511
macro avg	0.67	0.57	0.59	4,511
weighted avg	0.94	0.96	0.95	4,511

○ <표 6-18>은 위반과 미위반의 예측 정확도를 정리한 것이며 미위반에 대한 f1스코어는 0.98임

○ 이에 대한 설명으로 모형은 미위반에 대한 예측을 98% 정도의 정확성으로 예측함

○ 위반에 대한 f1스코어는 0.20임. 즉, 대구광역시의 단속원들이 조사를 나갈 시 9년 동안의 경험을 근거한 2018년 단속률은 3.9% 밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률은 20%까지 향상될 것으로 봄

7) 업체추천 결과(주요 6대 도시 통합)

<표 6-19> 주요 6대 도시 업체추천 모형별 f1스코어

모형	f1-score
KNN	0.082
로지스틱 회귀	0.152
의사결정나무	0.06
그래디언트 부스팅	0.157
에이다 부스팅	0.157
랜덤 포레스트	0.09
다층 퍼셉트론	0.118
서포트 벡터 머신	0.157

- <표 6-19>에서는 통합된 주요 6대 도시의 2018년 데이터를 각 모형으로 가동하였음. 가동한 모형 중 f1스코어가 가장 우수한 성능을 보인 모형은 서포트 벡터 머신, 에이다 부스팅, 그래디언트 부스팅임
- 모형 3개가 같은 점수를 보였으며 그중 서포트 벡터 머신을 사용함
  - 이유는 부스팅 모형들은 훈련데이터를 학습하고 예측할 때 맞추지 못한 오차의 데이터가 존재함. 이 오차의 데이터에 대해 가중치를 주어서 학습을 다시 하기에 다음번의 예측에는 오차가 줄어들. 이런 과정을 늘려나가기에 훈련데이터에 과대적합되어 검증데이터에 대한 오차율이 높아지기 때문에 서포트 벡터 머신을 사용함

<표 6-20> 분류결과표(Confusion\_Matrix)

모형 \ 실제	미위반(0)	위반(1)	총합
미위반(0)	50,984	-	50,984
위반(1)	1,181	110	1,291
총합	52,165	110	

- <표 6-20>를 보면 실제 단속 결과는 행(row)으로 모형이 예측한 단속 결과는 열(column)로 표시되며 모형이 미위반이라고 예측한 개수는 총 52,165개임. 이 중에서 맞춘 것은 50,984개이고 틀린 개수는 1,181개임. 모형이 위반에 대해서는 110개만 예측함. 전체 실제 위반 데이터 개수 1,291개 중 110개만 예측하고 맞춘

<표 6-21> 분류결과표의 f1스코어

	정밀도	재현율	f1스코어	support
미위반(0)	0.98	1.00	0.99	50984
위반(1)	1.00	0.09	0.16	1291
micro avg	0.98	0.98	0.98	52275
macro avg	0.99	0.54	0.57	52275
weighted avg	0.98	0.98	0.97	52275

- <표 6-21>는 위반과 미위반의 예측 정확도를 정리한 것으로 미위반에 대한 f1스코어는 0.99임
- 이에 대한 설명으로 모형은 미위반에 대한 예측을 99% 정확성으로 예측함
- 위반에 대한 f1스코어는 0.157로써 반올림하면 0.16임. 즉, 주요 6대 도시의 단속원들이 조사를 나갈 시 9년 동안의 경험을 근거한 2018년 단속률은 2.4% 밖에 되지 않지만 머신러닝 모형을 활용한다면 단속률을 16%까지 향상될 것으로 봄



나. 머신러닝을 활용한 업체추천 결과 요약표

<표 6-22> 분류결과표의 f1스코어

	모형	실제 단속률	모형 단속률
서울	서포트 벡터 머신	2.3%	16%
대구	서포트 벡터 머신	3.9%	20%
대전	다중 퍼셉트론	2.6%	13%
광주	에이다 부스팅	2.4%	12%
인천	서포트 벡터 머신	2.2%	32%
부산	그래디언트 부스팅	2%	13%

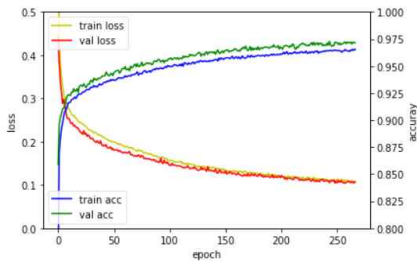
- <표 6-22>는 2018년의 주요 6대 도시별 실제 단속률과 모형의 단속률을 비교한 표임. 머신러닝을 사용할 때 서울의 경우에는 기존의 단속률 2.3%에서 모형 단속률 16%로 7배 정도 단속률이 향상됨. 대구는 기존의 단속률 3.9%에서 모형의 단속률 20%로 5배 정도 단속률이 향상됨. 대전은 기존의 단속률 2.6%에서 모형의 단속률 13%로 5배 정도 단속률이 향상됨. 광주는 기존의 단속률 2.4%에서 모형의 단속률 12%로 5배 정도 단속률이 향상됨. 인천은 기존의 단속률 2.2%에서 모형의 단속률 32%로 15배 정도 단속률이 향상됨. 부산은 기존의 단속률 2%에서 모형의 단속률 13%로 6배 이상 단속률이 향상됨. 주요 6대 도시의 데이터를 통합하여 머신러닝을 활용했을 경우 기존의 단속률 2.4%에서 모형의 단속률 16%로 6배 정도 단속률이 향상됨
  - 모형에 따라 점수는 각각 다르게 나왔으나 단속률 자체는 기존 단속률에 의한 것보다 더욱 높게 나왔음. 이러한 이유는 예측 정확도를 높여주는 변수들을 추가하였고 그 중 '재위반 여부' 변수가 위반자들에 대한 보다 효과적인 분류에 가장 큰 영향을 미쳤을 것으로 판단됨
- 132
- 향후 더욱 높은 예측 정확도를 위한 머신러닝의 모형을 위해 분석 예측에 효과적인 데이터를 활용할 것임. 현장의 공무원들의 노하우에 근거하고 분석에 용이할 수 있는 데이터를 연구하여 사용하고 분석 예측 과정을 최적화해주는 모형의 설계를 하여 예측 정확도를 높일 예정

## 2 단속 지역추천

### □ CNN을 활용한 지역추천 결과

- 훈련 진행 그래프와 정확도의 지표로 사용되는 f1스코어를 기준으로 모델을 평가함

<그림 6-1> Convolution Layer 훈련 그래프



- <그림 6-1>은 학습의 진행도를 나타내며, 날카롭게 뛰는 부분이 없는 부드러운 곡선이 학습의 진행상태에 문제가 없음을 보여줌

<표 6-23> 범주 판단 개수

	0등급	1등급	2등급	3등급	4등급	5등급	6등급	종합
실제	10,544	554	168	11	9	-	-	11,286
예측	10,539	463	139	94	17	25	9	11,286

- <표 6-23>를 보면 실제 값과 예측값이 매우 유사한 것을 볼 수 있음

<표 6-24> 분류결과표(Confusion\_Matrix)

	0등급	1등급	2등급	3등급	4등급	5등급	6등급	종합
0등급	10,444	59	35	1	-	5	-	10,544
1등급	89	72	351	41	1	-	-	554
2등급	6	8	77	44	15	18	-	168
3등급	-	-	-	8	1	2	-	11
4등급	-	-	-	-	-	-	9	9
5등급	-	-	-	-	-	-	-	-
6등급	-	-	-	-	-	-	-	-

- <표 6-24>을 통해 실제값을 기준으로 모델이 예측값이 어떻게 나왔는지를 확인할 수 있음. 실제 단속 결과는 행(row)으로 모델이 예측한 단속 결과는 열(column)로 표시됨

<표 6-25> 분류결과표의 f1스코어

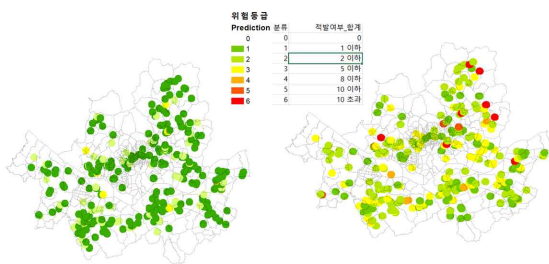
	정확도	재현률	f1스코어	support
0등급	0.99	0.99	0.99	10,544
1등급	0.52	0.13	0.21	554
2등급	0.17	0.46	0.24	168
3등급	0.09	0.73	0.15	11
4등급	0.00	0.00	0.00	9
5등급	-	-	-	0
6등급	-	-	-	0
accuracy	-	-	0.94	11,286
macro avg	0.25	0.33	0.23	11,286

- <표 6-25>를 보면 support는 각 등급별 샘플의 수를 나타낸 것. f1스코어는 모델의 각 등급별 예측률을 보여줌. accuracy는 전체 샘플에 대해 맞춘 비율임. 이 수치가 높은 이유는 0등급 샘플이 수가 많기 때문임. macro avg는 f1스코어의 평균을 나타냄. 이 수치가 0.23으로 나온 이유는 0~6등급까지의 모든 f1스코어를 평균으로 나타내

기 때문임. 따라서 실제 0등급에 대해서 0.99의 예측률을 보였으며 이는 위반가능성이 없는 지역에 대해서는 매우 높은 정확도를 보여 가지 않아도 될 지역은 잘 분류해주는 것으로 나타남. 0등급을 제외하고 위반가능성이 있는 지역에 대한 점수는 1~4등급까지만 표기가 되어 있는데 그 이유는 모형은 0~6등급까지 예측을 했으나 실제값은 5~6등급이 없기 때문임

- 1~4등급의 예측률은 0등급에 비해서 매우 낮게 나왔음. 1등급은 21%의 예측률을 보여주었고 2등급은 24%의 예측률, 3등급은 0.15% 예측률, 4등급은 0%의 예측률을 보여주었음
- 따라서 이 모형의 위험지역에 대한 최종 예측률은 각 등급의 예측률을 평균 낸 값으로 계산하여 31.8%라고 할 수 있음. 그러므로 단속원들이 이 모형을 사용하면 지역 선정에 대해서는 예측률을 32%까지 향상할 수 있다고 봄
- <표 6-24>를 보면 이 모형은 데이터를 과대 추정하여 5등급과 6등급을 표시하였음. 다음 그림은 모형에 예측한 지역추천 결과임

<그림 6-2> 왼쪽 그래프: 18년 실제값, 오른쪽 그래프: 18년 예측값



- 요약하면 18년도 서울지역을 대상으로 모형이 예측한 결과 과대 추정을 하였으며 1등급 이상 높게 잡은 것으로 나타났음. 결과적으로 지역 선정 예측률은 32%임
- 데이터를 수치적 형태가 아닌 이미지 형태를 사용한다면 CNN모형이 본 결과치보다 더 나은 예측 정확도를 나타낼 것이라 기대됨

## VII 원산지 표시 단속추천 시스템 활용 및 구축방안

### 1. 단속추천 시스템 활용방안

- 단속추천 시스템은 웹 기반으로 만들어져 PC, 태블릿 PC, 스마트폰 등 인터넷 브라우저로 접속이 가능한 모든 기기에서 활용할 수 있음
- 또한, 농관원의 조사원이 사용하는 스마트폰 기종에 맞춰 안드로이드 기반에서 사용하기 적합하게 제작함

#### 가. 업체·지역추천 시스템 활용

- <그림 7-1>은 인공지능 알고리즘의 위반 업체에 대한 예측 결과를 지도 위에 표시한 것으로 지도의 크기는 자유자재로 설정 가능함
- 추후 아이콘을 터치·클릭하여 상세정보를 보여주는 등의 편의성 기능 추가 가능함

<그림 7-1> 시안 프로그램 업체추천 현황 (영등포구)



138

- 추천업체는 <그림 7-2>와 같이 테이블 형태로 추천업체의 상호명, 업종, 주소 등 상세정보를 조회할 수 있음
- 추후 프로그램 개발 시 상호명을 클릭·터치하여 해당 위치로 이동하는 등 조사 편의성을 위해 추가적인 기능 탑재 가능함
- 지도상 추천업체 조회에 맞추어 추천업체 정보를 필터링하는 기능 탑재 가능함

<그림 7-2> 시안 프로그램추천 업체 정보 (영등포구)

상호명	업종	주소
소문난부지파지국밥	일반음식점	서울특별시 영등포구 국제금융로8길 25 (여의도동)
기름왕면도	슈퍼	서울특별시 영등포구 디지털로37길 29, 1층 (대림동)
대보름	일반음식점	서울특별시 영등포구 국회대로70길 22 (여의도동)
마더인디아	일반음식점	서울특별시 영등포구 국제금융로 86, 지하층 102호 (여의도동, 롯데캐슬 아이비)

- <그림 7-3>은 적발 위험지역을 지도상에 표시한 것임, 앞서 언급된 대로 적발 위험지역은 0에서 6까지로 정의됨
- 예시의 경우 문래동 1가에 위험등급 6의 지역이 신길로 33길 옆에는 위험도 2의 지역이 있는 것으로 나타남

<그림 7-3> 시안 프로그램 지역추천 현황 (영등포구)



139

### 나. 맵 커스터마이징 기능

- 본 시스템은 조사 지역·업체의 추천 뿐 아니라 분석에 활용된 주요 변수들을 지도상에 표시하여 조사원이 개인적인 노하우를 적용해 조사를 진행할 수 있음
  - 예를 들어, 조사원이 급식과 관련된 지역들을 조회하고 싶다면 주요 변수들 중 위락급식업체, 집단급식소, 초·중·고등학교 등의 변수를 선택하여 가장 밀집도가 높은 지역부터 조사를 실시할 수 있음
  - 현재는 단순히 클러스터링 기법을 사용하여 표현하였으나 히트맵, 그리드 등 다양한 형태로 표시 가능함

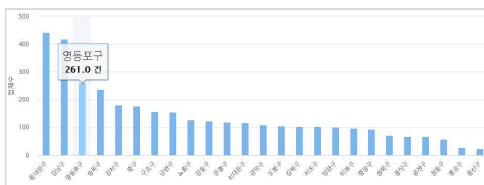
<그림 7-4> 시안 프로그램 클러스터링 기능



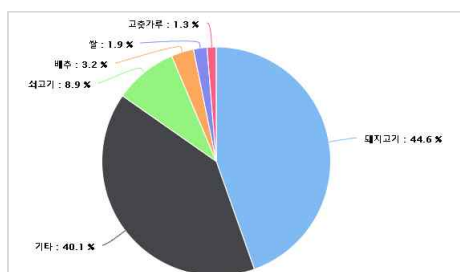
## 다. 실시간으로 활용 가능한 대시보드 기능

- 추천 시스템을 통해 추천된 예상업체의 현황을 실시간으로 파악할 수 있음
- 농관원의 데이터를 체크리스트 형태로 선택하여 실시간 그래프화인 이 가능하도록 구현 가능함
- 이러한 실시간 대시보드를 통해 데이터에 근거한 단속 업체 및 지역 선정에 도움을 줄 수 있음
- <그림 7-5>, <그림 7-6>에 표시된 막대 및 원그래프 외에 데이터 설명에 효과적인 다른 그래프도 추가할 계획임

<그림 7-5> 시안 프로그램 대시보드 - 서울시 구별 추천업체 현황



<그림 7-6> 시안 프로그램 대시보드 - 서울시 추천업체  
취급품목 현황



### 가. 현황

- 하드웨어 성능 한계: 본 프로젝트에서 사용되는 머신러닝과 딥러닝 모두 상당한 크기의 하드웨어 용량과 연산능력을 요구함
  - 특히 딥러닝(CNN)의 경우 부족한 하드웨어 역량으로 인해 분석에 많은 어려움을 겪었으며 기존 시스템으로 지역추천 알고리즘의 고도화나 전국단위 분석에는 무리가 있음
- 기존 방식에서 더 향상된 분석방법을 활용하여 지역추천 시스템을 가동하기 위해서는 높은 사양의 하드웨어를 필요로 함
- 서버운영: 시스템 시범운행을 위한 서버 필요
  - 시연 당시 사용한 서버 프로그램(ngrok<sup>1)</sup>)은 무료 시험용으로 실사용에 많은 제한이 있음
  - 시연 당시 ngrok을 활용하여 노트북에 로컬서버를 개설하여 사용
  - ngrok을 이용한 로컬서버 사용 시 접속 인원이 제한됨 (1~2명)
  - 서버 용량의 제한으로 인해 시연 당시 1대의 단말기에서도 지형정보의 표현이나 데이터의 시각화 등에 있어 한계점이 존재하는 것을 확인함
  - 클라우드 서버 등 별도의 서버가 필요함
- 서버 문제의 해결을 위해 국가에서 제공하는 클라우드 서비스 활용을 우선적으로 고려함. 용량이나 시기 등을 고려할 때 차선책으로 KT, Naver 등 국내 기관의 유상 서비스 활용 역시 고려함

1) ngrok은 NAT와 방화벽 뒤에 있는 로컬서버를 안전한 터널을 통해 공개 인터넷에 노출시켜 주는 도구, NAT는 네트워크 자원 절약 및 보안 강화를 위해 IP 주소 변환 시 사용하는 IP 주소 변환 기술

### 나. 단속추천 시스템 개발 개요(2020년도 계획)

#### 1) 2019년 데이터를 이용한 추천 알고리즘 최종 검정

- 2020년 원산지 단속시스템 프로토타입의 구현 및 시범운영을 위해 2019년도 데이터를 기반으로 모델의 검정 및 튜닝 실시
- 인공지능 알고리즘을 이용한 분석은 기본적으로 과거의 데이터를 이용해 패턴을 찾아내는 학습과정과 학습결과와 정확도를 분석하여 검정하는 두 개의 과정을 거침
  - 테스트를 위해 2009~2018년까지의 데이터를 학습 데이터로 지정하여 컴퓨터에 학습시킴
  - 2019년도 데이터를 검정 데이터로 지정하여 학습결과를 이용해 단속추천업체와 추천지역을 분석함
  - 업체추천은 머신러닝 알고리즘으로 지역추천은 CNN 알고리즘으로 분석이 진행됨

#### 2) 지역 추천 알고리즘의 고도화

- 지역 추천의 경우 컴퓨터 용량 부족으로 인해 진행하지 못했던 지역단위 이미지를 활용한 CNN 기법을 활용하여 지역 추천의 고도화를 진행할 예정
- 고도화된 알고리즘으로 전국단위로 분석범위 확장

#### 3) 단속서비스 프로토타입 고도화

- 기존에 서버문제로 인해 시연에 한계가 있었던 지도상의 시각화 문제를 해결하고 가용변수 추가 예정
- 대시보드를 필요한 형태로 사용자가 직접 조정·시각화 할 수 있도록 고도화 예정
- 논의를 통한 기타 부가기능 탑재

#### 4) 프로토타입 시범운영

- 구현한 프로토타입을 활용하여 시범 지역 내 현장에서 가동함으로써 현장에서의 적용가능성을 확인하는 것으로 진행 예정이며, 시범운영 기간은 2020년 말까지로 설정함
- 시범운영 대상 지역은 서울특별시로 지정하며, 약 2~30개의 단말기를 가동할 수 있을 것으로 예상됨(예산에 따라 조정 가능)
- 인공지능 시스템의 특성상 시행된 후에도 상당히 많은 튜닝 및 업데이트가 필요할 것으로 예상됨

#### 5) 시스템 실제 구현 및 운영(2021년 이후 계획)

- 현장에서의 시범운영 이후 2021년에 상용화를 시작하는 것을 목표로 진행할 예정이며 <표 7-1> 내용을 참조 요망
- 기본적으로 APP으로의 구현을 목표로 함
  - APP 형태로의 구현을 위해서는 별도의 예산 배정이 필요함
- 구현한 프로그램을 농림축산식품부 플랫폼에 탑재하여 모든 조사원들이 활용할 수 있도록 배포함
  - 농림축산식품부 내 다른 시스템과도 연동하여 편의성 및 정확도 증가
- 지속적이고 주기적인 튜닝·업데이트 및 관리 시행

<표 7-1> 단계별 시스템 구축 계획(예시)

기간	내용	비고
단기 (2020년)	- '19년 데이터를 활용한 시스템 검증 - 단속지역 추천서비스 고도화 - 웹과 앱 형태로 시스템 구현 - 대시보드를 통한 시각화 기능 강화 - 구현된 프로그램의 현장검증 실시 - 우선적으로 서울지역 현장검증 및 피드백을 통한 모형 튜닝(시범 지역은 상호 협의하여 결정 예정)	- 농관원 데이터 보완 필요 - 고성능 하드웨어 필요 - 서버 구축 필요 - 농관원 · 비리더 간 TIF 구축
중기 (2021년)	- 시스템 스펙의 확장(순차적으로 전국 단위 확장) - 시스템 구축 예산 확보	
장기 2022년 이후	- 시스템 구축 - 농림축산식품부 플랫폼 탑재 및 운영 - 지속적인 튜닝 및 운영	-

(위 내용은 현지 사정 및 농관원과 비리더의 상호 협의에 따라 바뀔 수 있음)

다. 단속추천 시스템 배포 및 관리

- 단속추천 시스템 관련하여 클라우드 서버 시스템으로 적극적으로 고려하고 있는 방법은 공공기관에서 제공하는 서버를 임대하여 진행할 예정이고 현재 국가가 공공기관을 통해 제공하는 기본 사양은 <표 7-2>와 같으며 지원 규모는 GPU 서버 1식을 제공함
- 이 클라우드 서비스를 받기 위해 12월에 지원 요청을 할 예정이며 임대가 확정될 경우 이를 활용하여 추가적인 개발을 진행할 예정임 (클라우드 서비스 사용 시점은 2월 말부터 가능할 것으로 예상)
- <표 7-2>과 같이 국가에서 제공하는 클라우드 서비스는 무료임

<표 7-2> 국가에서 제공하는 클라우드 서비스 예시

기본제공 자원	개발 프레임워크	인프라	지원서비스
CPU : 16core, 21GHz 이상	TensorFlow	데이터 관리·이관	개발환경 제공
GPU : 20TFLOPS 이상	PyTorch	백업 및 복구	시스템 사용교육
메모리 : 128GB 이상	Keras	보안관리	시스템 문의대응
HDD : 2TB 이상	Caffe	장애처리	...
1Gbps 회선 이상	...	...	...

※ 상세한 지원 규모는 향후 변경될 수 있음

- 또한, 이와는 별도로 차선책으로 유상 임대 방법으로 Kt, Naver 등이 국내 기업이 운영하는 클라우드 서비스를 임대하여 진행하는 것을 고려 중이며 이와 관련한 임대 조건과 예상 비용은 농관원과 협의하여 진행 예정
- 또 다른 방법은 해외 기업 Amazon사의 AWS를 사용하는 방법이 있음. <그림 7-7>은 클라우드 서버 사용 방안 중 하나인 AWS(Amazon Web Services)의 프리 티어 오퍼 유형으로 프리 티어를 이용해 60가지가 넘는 제품을 체험하고 AWS에 구축할 수 있음

146

- 프리 티어 오퍼는 AWS 신규 고객에게만 제공되며, AWS 가입일로부터 12개월 동안 제품 및 서비스를 무료로 사용할 수 있음<sup>1)</sup>
- 무료 사용 기간이 만료되거나 애플리케이션 사용량이 프리티어 범위2를 초과할 경우에는 사용량에 따라 표준 서비스 요금이 부과될 수 있음

<그림 7-7> 세 가지의 프리 티어 오퍼 유형



출처: <https://aws.amazon.com/ko/free/>

- 우선적으로 국가에서 제공하는 클라우드 서버 서비스를 고려하여 진행하며 차선책으로 Kt, Naver 등 국내 기관을 통한 유상 서비스를 고려하고 있음

1) <프리티어 세부 정보>, <Amazon Web Services>, <https://aws.amazon.com/ko/free/>, 검색일: 2019.11.28  
 2) 프리티어 범위는 12개월간 무료로 이용할 수 있는 AWS의 사용량을 의미. 서버구축을 위해 사용하는 AmazonEC2의 경우 월별 750시간의 인스턴스 사용량을 주고 이 범위를 초과할 경우 사용량에 따라 표준서비스 요금을 지불



## 라. 사용 프로그램

### 1) 지형정보 분석: QGIS와 ArcGIS

- 지형정보 분석에는 QGIS를 우선적으로 사용하며 추가적으로 고려하고 있는 것을 ArcGIS임
- QGIS는 오픈소스 이므로 향후 활용 가능성을 고려하는 경우 유료 서비스인 ArcGIS 보다는 QGIS를 활용하여 진행하는 것이 비용적인 측면에서 유리할 것으로 판단함

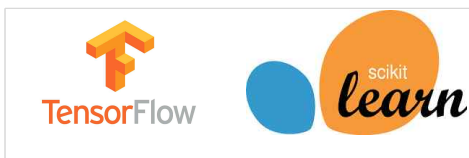
<그림 7-8> QGIS와 ArcGIS



### 2) 인공지능 알고리즘 분석: 텐서플로우(Tensorflow)와 사이킷런(Scikit-learn)

- 머신러닝과 딥러닝 등 인공지능 분석에는 텐서플로우와 사이킷런이 사용됨
  - 지역추천과 업체추천 모두 오픈소스 언어 Python의 라이브러리를 활용하여 구현하였음
  - 지역추천의 경우 TensorFlow 라이브러리, 업체추천의 경우 Scikit-learn 라이브러리를 활용하였음

<그림 7-9> 텐서플로우 및 사이킷런



### 3) 지도 구현: 폴리움(Folium)

- 지도 구현은 오픈소스 언어 Python의 Folium 라이브러리를 활용함
  - 폴리움(Folium)은 Open Street Map과 같은 지도데이터에 Leaflet.js를 이용하여 위치정보를 시각화하기 위한 라이브러리임

<그림 7-10> Folium



### 4) 프로그램의 웹 구현: 장고(Django)

- 웹 구현은 오픈소스 언어 Python에서 제공하는 웹 프레임워크 - 장고(Django)를 활용함
  - Python을 지원하는 웹 프레임워크 중 플라스크(Flask)를 사용하여 심플하게 표현할 수 있으나 사용자에게 반응형 웹을 제공하기 위해 장고를 활용함

<그림 7-11> Django



#### 5) 서버(Server): KT G-Cloud 서버

- 서울시 시범운행을 위한 웹 서비스에 Kt의 G-cloud를 사용할 계획임
- 클라우드 내부에 가상화된 VM을 생성하고 해당 서버에 Django로 구현한 추천 시스템을 삽입하여 서비스함

<그림 7-12> KT G-Cloud



#### 6) 데이터베이스: PostgreSQL

- PostgreSQL 상에 업체 정보를 입력하여 데이터베이스(DB)를 구현함

<그림 7-13> PostgreSQL



150

#### 7) 대시보드 구현: 하이차트(HighChart)

- 대시보드 구현하기 위해 HighChart의 API를 사용함
- 파제의 추가적인 진행을 위해서는 상업적 라이선스로 HighChart JS 버전의 구매가 필요함

<그림 7-14> High Chart



#### 8) 주소정제 기능

- 본 연구에서는 주소 정제를 위해서 오픈소스 API를 활용하여 지오코딩(geocoding)을 실시하여 분석에 활용함
- 하지만 변환할 자료가 전국단위로 매우 많은데 반해 처리속도가 사용이 힘들 정도로 느리고 많은 수의 자료가 좌표값을 내지 않거나 오차가 발생하는 등 많은 문제점이 발생함
- 이를 해결하기 위해 서울시가 사용하는 주소 정제 솔루션을 구매하여 진행할 예정임

1) 지오코딩(Geocoding): 입력한 주소를 좌표값으로 변환하여 지도에 점으로 찍어주는 작업

적용업무	사용 프로그램	비고
지형정보 분석	QGIS	-
지역추천	Tensorflow(파이썬)	-
업체추천	scikit-learn(파이썬)	-
지도 구현	Folium(파이썬)	-
웹 구현	Django(파이썬)	-
클라우드 서버	KT 클라우드 서버	150 / 월
데이터베이스	PostgreSQL	-
대시보드	Higchart JS	100
주소정제	주소정제 솔루션	300

※ 위의 프로그램은 예시이며 비용과 최종 프로그램은 협의하여 향후 변경될 수 있음

## VIII 요약 및 결론

### 1 요약

#### 1) 배경 및 데이터

- 이번 연구는 기존의 관행적 또는 신고 위주의 원산지 단속에서 과학적 데이터 분석 기반의 체계적이고 체계적인 단속으로의 패러다임 전환이 요구됨에 따라 원산지 관련 빅데이터 정보에 대한 종합분석을 통하여 다양한 조사·단속 패턴을 밝혀내어 향후 원산지 단속추천 시스템 개발을 위한 사전 기반 구축을 시도한 것에 의의가 있음
- 원산지 단속추천 및 선제 대응 서비스를 개발함에 있어 가장 먼저 ‘데이터 수집 및 처리’를 진행하였으며 데이터는 지역 및 환경변수, 업종 및 업체특성, 외부요인 변수 등을 수집하여 진행을 함
- 원산지 단속업무를 위한 한정된 자원의 효율적인 활용과 더욱 효과적인 단속, 계도 효과 창출 방안을 도출하기 위해 10개년의 농관원 단속 데이터를 이용하여 분석을 진행함

#### 2) 활용모형 및 분석과정

- 수집된 데이터는 통계 계산과 그래프를 위해 특화된 프로그래밍 언어이자 소프트웨어 환경인 R을 이용하여 전처리 과정 거침
- 수집된 데이터와 분석내용을 바탕으로 위반 업체 예측 모델링과 위반 지역 예측 모델링을 통해 단속추천 서비스를 개발하였으며 위반 업체 예측 모델링 관련 사용된 알고리즘으로는 KNN, 로지스틱 회귀, 의사결정나무, 랜덤 포레스트, 그래디언트 부스팅, 예이다 부스팅, SVM, 다층 퍼셉트론 등 다양한 머신러닝 기법을 사용하였으며 지역 단위 예측 모델링은 CNN을 사용함

- 서비스 개발에 이용된 분석 도구로는 데이터 전처리 및 분석을 위한 R, Python, Excel이 있고 모형구성에는 Tensorflow와 Scikit-learn을 활용하였으며 시스템 구현에는 django, folium, ArcGIS등을 사용함
- 단속 업체추천서비스를 위한 머신러닝 기법은 6대 도시에 대해 예측을 실시하였으나 단속 지역추천 서비스를 위한 CNN 기법은 6대 도시가 아닌 서울지역만 예측을 실시함
- 머신러닝의 결과는 다양한 모형을 활용하여 그 중 예측 정확도가 가장 높은 모형을 선택하여 사용하였음
- 주소 정제를 위해서 서울시가 사용하는 주소 정제 솔루션을 구매하여 진행함

### 3) 프로토타입 구축

- 분석결과를 바탕으로 웹브라우저와 스마트폰에서 사용 가능한 지도 기반의 추천서비스의 프로토타입을 구축하고 이를 통해 지도상에 위반 업체에 대한 정보와 분석에 활용된 주요 변수들의 정보를 제공함
- 농관원 데이터에 대한 실시간 그래프를 대시보드에 추가하여 데이터에 근거한 단속업체 및 지역 선정 판단에 도움을 줄 수 있음
- 본 분석을 통해 원산지 단속추천 시스템의 가능성은 확인이 됨

## 2 결론

- 6대 도시의 종합분석에 대한 시각화를 진행함으로써 6대 도시별의 특성을 파악할 수 있었으며 향후 이것을 웹이나 앱으로 구현함으로써 현장에서 사용할 수 있는 기틀을 마련
  - 조사·단속 상관관계 분석결과 원산지 표시 위반의 적발 건수 추세와는 관계없이 단속활동은 매년 유사한 형태로 이루어짐
  - 매년 비슷한 시기에 시즌별로 진행되는 정기단속의 경우 조사 취지와 맞지 않는 결과가 도출되고 있어 품목 조정 등 이에 대한 시정요구됨
  - 재위반 경향성 분석 결과 한번 적발된 업체가 다시 원산지 표시 위반을 하는 경향이 강하게 존재함을 알 수 있음
- 이번 연구를 통해 축적된 빅데이터를 종합적으로 분석하여 효율적인 조사방안을 도출하고 지도기반 위에서 신속 정확하게 조사대상을 추천받아 효율적으로 조사하는 선제적 사전 조사방법 개발 등을 통해 단속원들의 조사 효율성을 증대시킬 수 있는 기반을 마련
- 위반업체 추천서비스를 사용하는 경우 기존의 3%대의 적발률에서 12%대 이상의 적발률로 4배 이상의 효과가 있을 것으로 기대할 수 있음
  - 이를 금액으로 환산하는 경우 3%대인 적발금액이 136억여원에서 500억원 이상으로 상향될 것으로 기대됨
  - 2019년 데이터를 활용하여 검정과정을 거칠 필요가 있음
- 지역 추천의 경우 30~40%의 정확도를 보임
  - 고성능 하드웨어가 확보되면 이미지 데이터를 활용하여 분석을 시행하는 등 고도화를 통해 더 높은 정확도 도출이 가능할 것으로 예상됨
- 그 결과 위반업체와 지역추천 시스템을 활용 시 이전 보다 더 효율

적인 단속 활동을 할 수 있을 것으로 기대됨

- CNN을 활용한 단속 지역 추천 서비스의 경우 고성능의 하드웨어 확보와 이미지 데이터를 활용하여 지역 추천 서비스의 고도화를 통해 더 높은 정확도 도출을 할 필요성이 있음
- 데이터 과학에 입각한 추천 서비스의 타지역으로 확대 및 적발 고위험 지역 예측을 통한 선제적 단속추천 서비스 제공의 기틀 마련

### 3 제언

- 데이터 수집 방법 개선
  - 조사를 나가는 경우 어떤 품목을 조사를 하였는지 등 단속 업체별 모든 검사품목 기록 필요
  - 표시대상 업소의 사업자등록증 번호 등 다른 데이터와 연결할 수 있도록 해주는 Key 값을 가질 수 있는 데이터의 보완이 필요하며 이를 통한 조사 및 단속 대상 업체에 관련된 다양하고 정확한 정보를 이용한 분석 가능
  - 미위반 업소에 대한 정보를 보다 자세하게 표기
- 2019년 데이터를 통한 성능 테스트
  - 2019년도 데이터를 가지고 특정 도시 혹은 지역을 대상으로 하여 성능 테스트와 프로토타입을 구축하여 진행
  - 웹과 앱을 통한 시각화 진행
  - 2019년도 데이터를 기반으로 추천 서비스의 성능 테스트를 진행
  - 단속 지역 추천 서비스의 경우 이미지 데이터를 활용한 지역 추천 서비스의 고도화를 진행한 후에 진행
- 현장에의 적용
  - 위반 지역 추천 서비스의 경우 추가적인 개발을 위한 고성능의 하드웨어 확보가 필요함
  - Test Bed로서 한 도시나 지역을 지정하고 현장에 적용 및 현장의 견을 반영하여 고도화 하는 것을 제안함
  - 연구원들이 안정적으로 연구 개발을 할 수 있는 환경을 구축하고 이를 통해 Pilot 프로젝트를 위한 농관원과 빅리더 간의 TFT 구성이 필요
  - 현장의 의견이 주기적으로 반영될 수 있도록 하는 필드 테스트를 위한 프로토타입 개발을 통해 서비스를 고도화

- 앞서 설명한 순차적 단계를 통해 농관원은 단속추천 서비스 개발을 저비용으로 진행하면서 위험을 제거할 수 있는 장점이 있으며 빅리더 측에서는 개발된 알고리즘과 서비스를 특허 출원을 통해 농관원에 활용할 수 있는 장점이 있을 것
- 2019년 데이터를 통한 성능 테스트와 현장에의 적용을 하는 데 걸리는 시간은 약 1년 정도 소요될 것으로 예상되어지며 데이터 확보, 데이터 전처리, 알고리즘 및 서비스 개발 등에 약 7~8명의 인원이 필요할 것으로 예상됨
- 지자체 및 유관기관의 협조를 받아 향후 양질의 데이터를 얼마나 확보할 수 있느냐가 단속추천 서비스의 중요한 변수가 될 것이며 향후 알고리즘과 서비스 개발을 통해 고도화를 순차적으로 진행을 할 예정임
- 알고리즘 및 서비스 개발을 하는 데 있어 개인정보 관련한 사항은 사전에 충분히 검토하고 진행함으로써 이에 대한 향후 논란의 소지를 원천적으로 차단할 예정임
- 단속추천 시스템 개발을 위해 시중에서 용역을 통해 1년 기간으로 진행이 되는 경우 몇 억원대의 가격으로 진행이 되지만 특허 출원 및 활용 등 협력모델을 통해 순차적으로 농관원 현장의 의견을 반영하고 데이터를 확보하면서 2019년 데이터를 통한 성능 테스트와 현장에서의 필드 테스트를 순차적으로 진행을 하는 것이 개발의 리스크를 줄이는 좋은 방법이 될 것임