EMB00003ae06aad

201 학년도 제 학기

졸업작품/논문 최종보고서

제목 :

홍길동(학번)

(※팀원 이름 모두 기재)

201 년 월 일

지도교수: o o o 서명

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 계획(10) | 주제(20) | 개념(20) | 상세(30) | 보고서(20) | 총점(100) |
|  |  |  |  |  |  |

\* 지도교수가 평가결과 기재

○ 보고서 작성시 본 안내 텍스트상자 삭제 바랍니다.

○ 평가표는 수정하지 않습니다.

○ 본문 쪽번호는 필수입니다.

○ 남의 그림을 복사하여 사용하지 않습니다. 본인이 직접그림을 그립니다.

○ 모니터화면 갈무리는 가능한 사용 자제합니다.

○ 그림에 사용된 폰트는 본문 폰트와 일치해야 합니다.

○ 그림의 최대크기는 본문 쪽의 반으로 제한합니다.

○ 그림 및 표의 설명(캡션)은 필수입니다.

단, 표는 위에 그림은 아래에 번호와 함께 설명을 답니다.

○ 원천코드는 부록에 포함시킵니다.

○ 최종보고서 분량은 총 30페이지로 작성합니다.

평가 배점

계획 (10)

무리 없는 진행.

주제 (20)

신규성, 진보성, 창의성, 현실성

개념설계 (20)

주제에 명시된 바대로 필요한 기능이 모두 포함되었는가?

상세설계 (30)

계획에 명시된 바대로 결과물과 일치하는가?

각각의 practical constraints에 기술된 사항을 만족하는가?

보고서 (20)

짜임새 있는 정리 및 표현

* **요약**

2016년 11월부터 시작된 조류 인플루엔자로 인해 대한민국은 산란계의 떼죽음으로 전례 없는 계란 값 폭등을 겪었다. 대부분 음식에 자주 사용되는 계란 원가가 폭등하자 외식업계는 음식 가격을 인상하거나 계란 사용을 일시 중지하는 등 전국적인 큰 혼란이 빚어졌다. 본 논문에서는 데이터 분석 기법을 이용해 조류 인플루엔자 바이러스와 해당 지역의 기온의 상관 관계를 밝혀내고자 한다. 기온 외에도 습도, 강수량 등 외부 환경 요인을 늘릴 수 있으나 데이터 수집의 한계로 본 논문에서는 기온과의 관계만 파해지고자 한다. 구체적으로는 회귀분석 기법을 이용해 조류 독감과 외부 기온의 상관관계를 파악한 후 조류 인플루엔자 발생과의 연관성을 측정하고자 한다. 조류 인플루엔자 발생 데이터는 농림축산검역본부 사이트에서 공개 데이터로 얻을 수 있고 기온 데이터는 기상정보개방포털 사이트에서 데이터를 얻을 수 있다. 본 논문을 통해 (1) 조류 인플루엔자 바이러스와 기온의 상관관계 측정, (2) 측정을 통해 향후 연구 혹은 활용방안으로써 활용할 수 있을 것이다.

* **서론 (6페이지 이내)**

1. **제안 배경 및 필요성**

2016년 11월부터 발생하기 시작한 조류 인플루엔자는 대한민국 사회에 지대한 영향을 끼쳤다. 이전부터 조류 인플루엔자로 인한 피해는 있었으나 그 여파가 일반 소비자 식탁에까지 미친 건 이번이 처음이다. 조류 인플루엔자로 인한 산란계 수 급감으로 계란의 공급이 현저히 감소했고 계란 값은 며칠만에 두 배 이상으로 치솟았다. 계란을 이용한 음식이 많은 한국인의 밥상 특성상 그 영향은 이루 말할 수 없을 정도로 지대했다. 2017년 3월 기준 살처분된 가금류의 수는 3000만 마리 이상으로 기존 사육되는 가금류 수의 20%를 차지한다. (정빛나, 2017) 이렇듯 조류 인플루엔자로 인해 경제적, 사회적 손실이 매우 큰 상황이지만 문제는 여기서 끝나지 않았다는 점이다. 조류 독감으로 심각한 피해를 본 양계장 업주들은 이를 예방하기 위해 인체에 해롭다고 판정된 살충제를 가금류에 투여했다. 결국 인간에게 해를 끼칠 수 있는 살충제 사용이 언론에 적발되면서 소비자는 계란 사용을 더욱 기피하게 되었다. 닭은 조류 인플루엔자로 인해 대량으로 살처분 되었고, 계란은 팔리지 않으니 양계업자들은 그야말로 사면초가에 빠진 것이다.

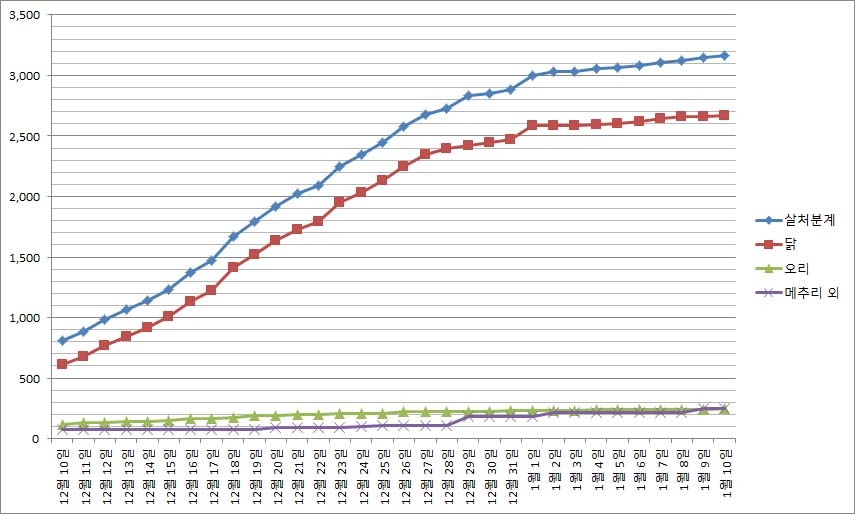


그림 1 2017년 1월10일까지의 살처분계 상황 그래프

이렇듯 조류 인플루엔자로 인한 일명 ‘계란 파동’은 11개월이 지난 지금에서야 서서히 안정을 찾기 시작했다. 하지만 살충제 파동으로 인한 계란에 대한 부정적인 시선은 여전히 존재하며 산란계가 급감한 양계업자들 또한 향후 산란계가 보충되기 전까지는 어려움이 지속될 것으로 보인다.

사회적으로나 경제적으로나 한국 사회에 심각한 피해를 준 조류 독감을 예방하는 방법으로는 방역 작업이 전부일 정도로 기술적 지원이 안 되고 있는 상황이다. 본 논문을 작성하면서 조류 인플루엔자에 예방과 관련된 다른 논문을 검색했지만 방역 작업의 효율성을 높이는 방법 외의 내용은 찾아볼 수 없었다.

하지만 UN(이하 유엔) 산하의 LEAD Company(이하 리드 컴퍼니) 사례를 통해 조류 인플루엔자를 빅데이터 분석을 통해 예방할 수 있을 것이라 생각한다. 리드 컴퍼니의 주 목적은 통신사, 보험사 등에서 얻은 고객 빅데이터로 감염병의 전파 경로를 예측해 이를 사전에 예방하는 것이다. 이 방법을 조류 인플루엔자에 대입하여 그 전파를 방지할 수 있다면 한국 사회에 이번 조류 독감 및 살충제 계란 파동과 같은 손실을 예방할 수 있을 것이다. 본 논문이 주장하는 필요성은 바로 조류 인플루엔자의 예상 발생 지점을 예측하고 이를 방지에 사회적 손실을 줄이하는 것에 있다.

1. **논문의 목표**

본 논문의 최종 목표는 조류 독감을 1차 발생에 그치게 하는 것이다. 사실 조류 인플루엔자는 철새에 의해 유입되므로 완전 예방은 불가능하다. 하지만 바이러스의 발생 지점을 예측해 막는다면 조류 독감으로 인한 피해를 대폭 줄일 수 있을 것이다. (김재만, 2015). 향후 습도, 날씨 등과 같은 다양한 데이터와 교차검증, 주성분 분석 기법 등 다양한 분석 기법을 이용해 더욱 정확한 발생 지점 예측 모델을 만들 수 있을 것이다.

1. **논문 전체 overview**

본 논문은 크게

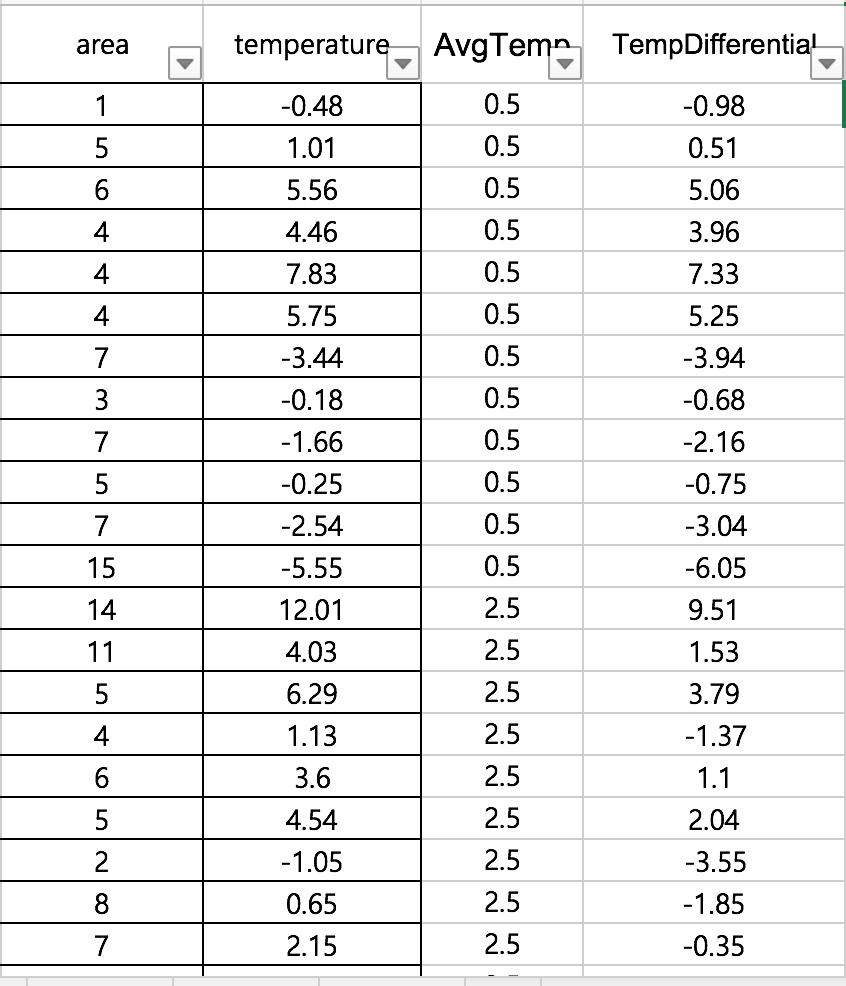
1. 데이터 수집, 생성 및 정리 상황
2. 정리된 데이터를 이용한 그래프 추출
3. 그래프와 수치 데이터를 이용해 회귀분석 기법을 실행하고, 기온과 조류 인플루엔자의 연관성 분석
4. 결론

이 4가지 부분으로 진행될 예정이다.

데이터 수집, 생성 및 정리 상황에서는, 조류 인플루엔자 신고 상황을 농림축산검역본부 사이트 (농림축산검역본부, 2017년)에서, 조류 인플루엔자 발생 지역의 기온 데이터 및 당시 한국의 평균 기온 데이터는 기상자료개발포털 (기상자료개방포털, 2017년) 사이트에서 데이터를 얻을 수 있었다. 그런 후 조류 인플루엔자 발생 지역과 날짜, 그리고 당시 해당 지역 기온 및 당시의 한국 평균 기온 데이터를 병합해 하나의 엑셀 파일로 정리하였다.

엑셀 파일은 열은 조류 인플루엔자 발생 구역(*area*), 당시 지역 기온(*temperature*), 당시 전국 평균 기온(*AvgTemp*) 그리고 기온과 전국 평균 기온의 차(*TempDifferential*)가 있다. 여기서 발생 구역(*area*)의 분류 기준은 제안 작품 소개에서 자세히 설명할 예정이다.

표 1 조류 인플루엔자 발생 지역, 당시 해당 지역 기온, 전국 평균 기온 그리고 기온차를 정리한엑셀 파일의 일부



데이터를 이용한 그래프 추출에서는, Python의 Pandas 등의 라이브러리를 이용해 그래프를 얻을 수 있었다. 그래프 추출 시작 단계에서 Python의 Altair 라이브러리를 이용해 그래프를 작성할 계획이었으나 Pandas의 사용법 및 사례가 더 많다는 것을 발견하여 Pandas를 이용하기로 최종 결정하였다. (PinkWink, 2015)

그렇게 그래프라는 시각적 데이터를 얻었고, 기온과 조류 인플루엔자 발생 지역 데이터 간의 연관성을 회귀분석 기법을 이용해 수치화하여 결론을 도출할 수 있었다. 시각적 데이터로는 자세한 연관성을 알아낼 수 없었지만, 회귀분석 기법에서 생성된 multiple R 값이 매우 높은 수치를 보여 기온과 조류 인플루엔자 간의 연관성이 높음을 알 수 있었다. Multiple R의 값과 관련된 글을 구현 및 결과분석에서 다룰 예정이다.

* **관련연구 (7페이지 이내)**
  1. 빅데이터와 전염병의 연관성

이전까지 빅데이터를 이용한 연구 분야는 주로 정치, 마켓팅, 신제품 기획 등에 사용되어 왔다. 빅데이터를 이용하여 전염병을 다루는 연구는 매우 적게 이루어졌으며 이 또한 큰 주목을 받지 못한 채 활용이 못 되고 있다. 하지만 2015년 발생한 메르스 사태 이후 사람들은 기존 언론에 대해 불신하기 시작했고 SNS를 통해서 전염병과 관련된 정보를 얻기 시작했다. (허지윤, 2017) 이와 같이 국내의 경우 메르스 사태를 통해 빅데이터를 전염병과 연관 짓기 시작했고 관련 연구도 진행이 되었지만 구체적으로 조류 독감의 확산과 빅데이터와 관련된 연구가 진행된 경우가 매우 드물다. 실제로 RISS와 NDSL에서 “조류 독감 빅데이터” 키워드로 검색할 시 RISS에서는 관련 논문 하나, NDSL에서는 아예 관련 논문이 검색되지 않는다. RISS에서 검색된 논문은 「빅데이터 기반 실시간 농축산 검역을 위한 확산경로 탐색과 역학조사 모델링 연구[1] 」으로 2010년 발생한 구제역과 농장 간 차량 이동을 연관 지어 특정 차량 방문 시 바이러스가 전염된다는 주제로 작성되어 있다. 이 논문에 의하면, 기존 역학 조사는 특정 농장 방문 차량 리스트를 추출하거나 특정 차량의 방문 농장 리스트를 추출하는 기법을 사용하고 있으나 이를 특정 농장 방문 차량을 유형별, 목적별, 기간별 등으로 분석하고 특정 차량의 방문 농장을 유형별, 가축종류별, 기간별, 발병이력여부별 등으로 분석하는, 보다 자세하고 빅데이터를 이용해 방대한 데이터를 통해 유의미한 데이터를 추출하는 것에 방점을 두고 기존 시스템을 개선하고자 하는 이론을 제시하였다. (김종현(KimJong, 2014)

* 1. 효과적인 논문 전파 방식

한편, 또다른 논문인 「페이스북 이용을 통한 정보획득과 사회자본이 질병 예방행동의도에 미치는 영향[2]」에서는 메르스로 인해 페이스북을 이용한 전염병 관련 소식을 접하는 사람들이 늘고 있다는 것을 총 324명의 설문조사를 통해 도출하였다. 여기서 사람들은 빅데이터를 이용해 전염병과 관련된 소식을 들으면 오히려 전염병 예방에 소흘해 진다고 주장하고 있다. SNS 상에는 검증되는 않은 이야기와 괴담이 난무하면서 사람들은 SNS에서 생성되는 소식을 귀담아 듣지 않는 것이다. 그러므로 본 논문의 주장을 전파하기 위해서는 더욱 다양한 데이터와 분석 기법으로 객관적인 수치를 제시한 후, 여러 학회를 통해 본 논문의 아이디어를 전문가들에게 전하여 설득하고 실천하게 만드는 것이 효과적일 것이다.

* 1. 전염병과 빅데이터의 연관성에 대한 정부의 태도 변화 필요

「사회안전을 위한 빅데이터 활용의 재난대응 정책[3]」에 의하면, 현대사회는 고도의 전산화, 정보화가 진행되었는데도 불구하고 전염병과 관련해서 정부의 대응이 미흡하다고 지적하고 있다. (정영철최익수,, 2016) 빅데이터를 이용하여 바이러스의 패턴을 분석하고 이를 효과적으로 예방할 수 있음에도 불구하고 정부는 바이러스가 창궐한 이후에야 적극적인 대응을 실시하고 이마저도 검역 강화 등 구시대적인 방식을 이용하고 있다. 메르스 사태와 2016년 11월 발생한 조류 인플루엔자 사태는, 그 시작은 비록 우연이었을지 몰라도 정부의 미흡한 대응 방식이 한 몫한 것이다. 만약 정부가 빅데이터를 이용해 조류 인플루엔자 등의 전염병을 적극적으로 예방한다면 서론에서 서술한 것과 같은 재앙이 발생하지 않았을 것이다. 그러므로 정부는 더이상 바이러스가 발생한 이후에야 검역 강화를 하는, 구시대적인 방식에서 벗어나 빅데이터와 통계학의 다양한 분석 기법을 이용해 전염병을 예방해야 할 것이다.

* 1. 해외에서 입증된 빅데이터를 이용한 전염병 예방 효과

「Study on Big Data in Public Health, Telemedine and Healthcare[4]」에 의하면, 빅데이터를 이용한 공중보건 관리는 권장 사항이라 하고 있다. (Osterreich, 2016) 그러면서 각국의 정책, 그리고 국민들의 관심 등이 필요하다 주장하고 있다. 「Big Data Analytics for Healthcare[5]」에서는 빅데이터와 공중보건을 어떻게 하면 효과적으로 융합할 수 있는지를 서술하고 있다. 예컨대 빅데이터로 환자의 상태를 상세히 분석하고 관련 사례를 통해 어떠한 전염병에 감염될지를 예측하여 사람들이 병원비로 사용하는 액수를 낮출 수 있다는 것이다. (Jimeng Sun, 2013) 「Big Data for Health[6]」에서는 빅데이터를 이용하여 전염병에 발생 가능성을 예측하고 관련 백신 연구를 선행하여 효과적인 전염병 전파 저지를 하는 것을 서술하고 있다. 또한 사람의 DNA, 인체 구성 등을 빅데이터로 정리해 감염되기 쉬운 바이러스를 분류해서 전염병을 예방하는 것에도 방점을 뒀다. (Javier Andreu-Perez, 2015)

* 1. 관련연구를 통한 고찰

상술한 관련연구를 참고하면서 본 논문은 본 연구의 방향, 전파 방식 등 다양한 방법을 알 수 있었다. 하지만 조류 독감과 빅데이터에 대한 분석 기법, 분석 사례 등을 연구한 논문은 찾을 수 없었고 그나마 빅데이터를 이용해 조류 독감 등 전염병을 예방하는 방법으로는 방역 작업의 세분화, 감염자의 이동 경로 추적 및 분석 등의 이동과 관련된 연구 밖에 없었다.

본 논문은 조류 독감의 전파에 영향을 미치는 요소가 방역 작업과 차량 이동 외에도 환경적 요인이 있을 것이라 확신하고 연구에 임했다. 그러므로 향후 데이터 분석 등을 통해 관련 연구가 미비했던 본 주제에 대한 결론을 도출하여 논문을 작성하고자 한다.

* 제안 작품 소개 (7페이지 이내)

실제로 구현에 대해 소개합니다. 자세하면 자세할수록 좋습니다. 이론적 배경과 시스템구성은 필수입니다.

■ 구현 및 결과분석 (7페이지 이내)

소정의 결과를 소개하고 결과에 대한 분석을 적습니다.

■ 결론 및 소감 (2페이지 이내)

결론을 맺습니다.

졸업작품/논문을 하면서 겪었던 소중한 경험들에 대해 적습니다. 가장 중요한 부분입니다.

■ 참고문헌 (1페이지 이내)

인용한 참고문헌입니다. 최소 20개 이면 영문 10개 국문 10개 이상입니다. 참고문헌 인용은 다음과 같이 합니다 [1].

[1] K.Thompson, G. J. Miller, and R. Wilder, "Wide-area Internet traffic patterns and characteristics (extended version)," IEEENetworkMagazine, November 1997.

[2] Network Simulator ns-2, http://www.isi.edu/nsnam/ns/.

[3]