1. **과제의 필요성**
   1. **Abstract**

작년 11월부터 시작된 조류독감으로 인해 대한민국은 산란계의 떼죽음으로 전례 없는 계란 값 폭등을 겪었다. 한국 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 기온 등 외부 환경과의 상관 관계를 밝혀내고자 한다. 더 자세히는 Regression Analysis 기법, Association Rule Learning 기법 등을 이용해 조류독감과 외부 환경의 상관관계를 파악한 후 조류 인플루엔자 바이러스의 경로를 예측하고자 한다. 조류독감 발생 데이터는 농림축산검역본부 사이트에서 공개 데이터로 얻을 수 있고 기온 등은 기상청 데이터를 통해 얻을 수 있다. 본 논문을 통해 조류독감의 예상 전파 지역을 미리 파악한 후 대처하여 조류독감 발생 빈도 감소를 기대해볼 수 있다.

* 1. **연구배경**

작년 11월부터 발생하기 시작한 조류독감은 대한민국에 엄청난 영향을 끼쳤다. 이전부터 조류 독감은 양계장에게는 반드시 경계해야할 대상이었지만 그 여파가 일반 소비자 식탁에까지 미친 건 이번이 처음이다. 계란 값은 며칠만에 두 배 이상 치솟았으며 계란을 이용한 음식이 많은 한국인의 밥상 특성상 그 영향은 이루 말할 수 없을 정도로 지대했다. 올해 3월 기준 살처분된 가금류의 수는 3,000만 마리 이상으로 이는 기존 사육되는 가금류 수의 20%를 차지한다. (정빛나, 2017) 문제는 여기서 끝나지 않는다. 조류 독감으로 인해 너무 많은 닭이 죽었고 이는 양계장 업주로 하여금 살충제를 사용하게 만드는 촉매제가 되었다. 결국 인간에게 해를 끼칠 수 있는 살충제 사용이 적발되면서 소비자는 계란 사용을 더욱 기피하게 되었다. 닭은 조류 독감으로 인해 대량으로 살처분 되었고 계란은 팔리지 않으니 양계업자들은 그야말로 사면초가에 빠진 것이다.

이렇듯 조류 독감의 영향은 10개월이 지난 지금까지도 소비자에게 영향을 끼치고 있다. 하지만 조류 독감에 대한 확실한 예방 조치는 나오지 않고 있으며 이는 가까운 미래에 이런 참사가 재발할 수 있다는 것을 의마한다.

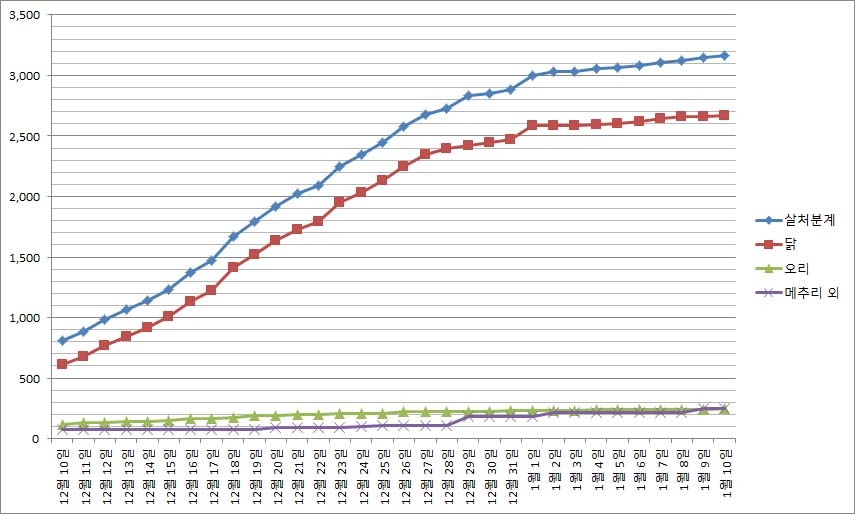


Figure 금년 1월10일까지의 살처분계 상황 그래프 (Jnana, 2017)

본 논문의 목적은 LEAD Company에서 사용 중인 빅데이터를 이용한 전염병 방지 기법을 조류 독감에 적용하여 조류 독감의 피해를 최소화하는 것이다. KT 등이 참여중인 LEAD Company 사업은 전 세계 수 많은 기업의 빅데이터를 이용해 전염병의 예상 발생지점 예측 및 전염 방지하는 것이다. (박지성, 2017) 물론 인간 전염병과는 달리 조류 독감은 철새에 의해 발생한다는 측면에서 완전한 예방은 불가능하다. 하지만 기존 발생 지역과 기온 등의 주변 환경간 관계를 분석하여 일정한 연관성이 나온다면, 조류 독감 최초 발생 후 차기 전염 지역을 예측할 수 있을 것이며 이는 조류 독감 전파 방지에 큰 도움을 줄 것으로 예상된다.

1. **선행연구 및 기술현황**
   1. **선행연구**

조류 독감의 이전 발생 경로를 통해 예상 전파 경로를 예측하는 선행연구는 거의 이뤄지지 않았다. 정부 차원에서의 예방 차원으로는 방역작업이 전부였으며 이 마저도 완전한 예방을 하지 못해 작년 11월부터 시작된 독감과 같은 심각한 상황을 발생시켰다.

조류 독감과는 별개로 데이터 분석 분야에서는 수많은 분석 기법이 쓰이고 있다. 거의 모든 산업 분야에서 빅데이터가 점점 늘어나고 있고 수집된 데이터를 원하는 형태로 정제하기 분석하기 위해 필요한 데이터 분석 기법은 나날이 많아지고 있으며 이는 빅데이터 성능에 결정적인 영향을 끼친다. 예컨대 아무리 데이터를 많이 수집하고 정제를 잘 한다 해도 분석 능력이 떨어진다면 무의미한 결론을 얻을 수 있으며 이는 빅데이터의 사용 의미 자체를 희석시키는 결과를 낳을 수 있다.

빅데이터 분석을 위해서 가장 먼저 선행되어야 하는 것은 분석하는 데이터 분야에 대한 이해이다. 분석자가 그 분야에 대해 무지하면 무엇을 원하는지, 목표는 무엇인지 알 수 없으며 데이터가 무엇을 의미하는지 알 수가 없다. 그러므로 데이터를 분석하기 전 데이터와 관련된 분야 공부는 필수이다.

* 1. **기술현황** (옥수별(samsjang), 2016)

다음과 같은 데이터 분석 관련 기술들이 활용되고 있다.

* + 1. **회귀분석(Regression Analysis)**

회귀분석 기법은 데이터 분석 관련 기술 중 하나로 어떤 변수가 다른 변수에의하여 설명된다고 보고, 그 함수 관계를 조사하는 통계학적 기법이다. 말시 말해, 변수 간의 상호 관련성을 찾아내고, 변수들의 변화로부터 다른 변수들의 변화를 예측하는 기법을 말한다. 변수 간 상호 관련성 분석 기법에는 상관 분석과 회귀 분석이 있는데 전자는 기본 변수와 목적 변수의 역할이 바뀔 수 있을 때나 역할의 구분이 없을 때 사용할 수 있으며, 후자는 두 변수의 관계가 분명한 방향이 있을 때 사용할 수 있다. 데이터 분석 기법에서는 주로 후자를 사용하는 경우가 많다.

* + 1. **연관규칙학습(Association Rule Learning)**

연관규칙학습이란 상품, 서비스 간의 관계에서 유용한 규칙을 찾아내는 기법으로 간단히는 If A then B로 표현할 수 있다. 특정 문제에 대한 예측 값이나 특정 목적에 따라 분류하는 것이 아니라, 상품이나 서비스의 거래 기록 같은 데이터에서 연관성 정도를 측정하여 관성이 많은 데이터를 그룹화하는 클러스터링의 일종이다.

본 졸업논문에서 얻고자 하는 결론은 조류 독감 발생 지역과 기온 등 주변 환경의 상관 관계에 대한 자료이고 이 데이터를 분석해야 한다. 이 두 데이터 간의 데이터에서 연관성 정도를 측정할 수 있을 것으로 기대되어 해당 기법을 이용하면 결론을 얻을 수 있을 것이라 판단된다.

* + 1. **유전알고리즘(Genetic Algorithms)**

유전알고리즘이란 자연계의 유전학에 기반한 병렬적, 전역적인 탐색 알고리즘으로, 다윈의 적자생존 이론을 기본으로 한다. 문제의 가능한 해를 정해진 형태의 자료구조로 만들고 점차적으로 변형하여 더 좋은 해를 만든다. 이때, 해를 나타내는 자료구조는 유전자, 변형에 의해 더 좋은 해를 만드는 과정을 진화로 볼 수 있다.

해당 논문 주제에서는 주변 환경을 점차적으로 변형시켜 더 좋은 연관 관계를 만들 수 있을 것으로 기대하고 있다.

1. **작품/논문 전체 진행계획 및 구성**

**[표 1] 논문 월별 진행계획**

|  |  |  |  |
| --- | --- | --- | --- |
| **월별**  **내용** | **9** | **10** | **11** |
| **제안서 작성 및  사전 조사** |  |  |  |
| **관련 기술 조사** |  |  |  |
| **조류 독감 관련 데이터 수집** |  |  |  |
| **기온과 주변 환경과 관련된 데이터 수집** |  |  |  |
| **데이터간 상관 관계 분석** |  |  |  |
| **중간 보고서 작성** |  |  |  |
| **결론 도출** |  |  |  |
| **최종 보고서 작성** |  |  |  |

본 졸업논문에서 제안하는 주제는 LEAD Company에서 전염병에 관한 데이터 분석 및 예측 기법을 조류 독감에 적용하고자 한다. 조류 독감의 발생지와 주변 환경에 따른 전파 경로를 분석하고 연관성을 도출하여 조류 독감의 전파를 보다 효율적으로 예방하고자 한다.

* 1. **제안서 작성 및 주제 관련 조사**

졸업 논문을 진행하기 전 제안서 작성 및 작품 진행에 관련된 사전 조사를 한다. 추후에 자세하게 진행할 조류 독감 발생 데이터 수집, 기온 등 지역 환경 데이터 수집, 데이터 분석 비법 등에 대해 사전 조사를 진행한다.

* 1. **관련 기술 조사**

관련 기술 조사는 사전 조사 시에 다룬 내용을 바탕으로 각각의 항목들에 대해 세부적인 조사 및 연구를 진행한다.

* 1. **조류 독감 관련 데이터 수집**

국내 조류 독감 발생 신고와 관련된 데이터는 매우 자세히 저장되어 있고, 원하는 사용자 모두가 수집할 수 있도록 농림축산검역본부에서 서비스를 제공하고 있다.

* 1. **기온과 주변 환경과 관련된 데이터 수집**

국내 지역 기온과 강수량 등의 환경 관련 데이터는 기상청 빅데이터 센터에 저장되어 있고 원하는 사용자 모두가 수집할 수 있다.

* 1. **데이터간 상관 관계 분석**

조류 독감과 기온 등 지역 환경 데이터를 바탕으로 데이터 분석을 실시해 상관 관계를 도출하고 파이썬 언어로 정리한다.

* 1. **결론 도출**

졸업 논문 진행 결과에 대해 결론을 도출하며 futurework를 제안한다.

1. **기대효과 및 개선방향**
   1. **기대효과**

본 논문을 통해 기대되는 효과는 바로 조류 독감의 전파 예방, 장기적으로는 국가적인 경제적 손실 및 사회적 혼란을 막기 위함이다. 서론에서 서술하였듯이 현재 조류 독감의 예방책으로는 방역 작업밖에 없으며 이는 완전한 전파 예방을 하지 못하고 있다. 하지만 조류 인플루엔자와 환경의 상관 관계를 밝혀낸다면 조류 독감 예상 전파 지역을 알아낼 수 있으며 해당 지역의 방역을 각별히 강화한다면 조류 독감의 효과적인 예방을 기대할 수 있을 것이다.

예컨대 LEAD Company에서 빅데이터를 이용해 전염병 전파를 예방하고 있듯이 국내 조류 독감의 데이터와 환경 데이터를 참조하여 이의 상관 관계를 알아내 향후 조류 독감 재발 시 예상 전파 지역에서 바이러스 전파를 원천 차단해 손실을 막는 것이다.

* 1. **개선방향**

본 작품의 개선 방향으로는 지역 환경 데이터를 더욱 세분화해 기온, 강수량뿐만 아니라 습도, 미세먼지 농도 등 모든 데이터와 조류 독감관의 상관 관계를 분석해 더욱 자세한 조류 독감 전파 예측이 가능한 것이다. 이렇게 된다면 조류 독감의 효과적인 방역 작업으로 경제적 손실과 사회적 혼란을 더욱 잘 막을 수 있을 것이다.

1. **기타**
   1. **진행계획**

진행계획은 아래 [표 2]에 나타내었다.

**[표 2] 진행계획**

|  |  |  |  |
| --- | --- | --- | --- |
| **월별**  **내용** | **9** | **10** | **11** |
| **연구노트 작성** |  | | |
| **작품 제안서**  **서약서** |  |  |  |
| **작품 중간보고서** |  |  |  |
| **작품 최종보고서** |  |  |  |
| **작품 발표회** |  |  |  |

# **6 참고문헌**

박지성. (2017년 1월 11일). “KT, UNGC 리드컴퍼니 가입···빅데이터 활용 감염병예방 세계전파”. 전자신문 : http://www.etnews.com/20170111000255에서 검색됨

옥수별(samsjang). (2016년 8월 17일). “빅데이터 사례 연구 - 데이터 분석 기법”. 네이버 블로그: http://sams.epaiai.com/220788857564에서 검색됨

정빛나. (2017년 1월 5일). “AI 재앙 언제 끝나나…"피해규모 1조원"”. 연합뉴스: http://www.yonhapnews.co.kr/bulletin/2017/01/04/0200000000AKR20170104182700030.HTML에서 검색됨