Analyzing Environmental Factors for Foodborne Illness and Recall Prediction

Project Proposal & Statement of Work

Project Manager: Janessa Davis

Alex Alonso

Nicholas Filiberti

Khusniddin Abdullaev

Xiunan Song

Mentors: Dr. Kristen Pogreba Brown & Dr. Erika Austhof

Date: March 13, 2024

Revision History Table		Template Date 6/15/2022
Version	Summary of Changes	Date
0.1	*Added Title Page and Executive Summary*	*01/17/24*
0.2	*Added Lit Review and Research Project Deliverables*	*01/24/24*
0.3	*Added Project Timeline and Ethics*	*02/05/24*
0.4	*Updating Mentor, Authors, and Approvals*	*02/12/24*
0.5	*Editing and updating based on peer feedback*	*02/19/24*
0.6	*Changing project topic and editing executive summary and lit review based on mentor feedback*	*03/11/24*
0.7	*Final revised proposal after updated research and mentor feedback*	*3/13/24*

Contents

1. Executive Summary

The Executive Summary was written by Janessa Davis, Khusniddin Abdullaev, and Xiunan Song.

Our research project is a statistical and machine learning analysis that investigates regional environmental factors such as weather changes and temperature patterns in relation to food-borne recalls and outbreaks that have been seen over the years. This study aims to investigate the relationship between the climate changes for major agricultural production regions in the United States and the likelihood of food-borne outbreaks such as *Listeria sp* and *Salmonella*.

As organizations such as the FDA and CDC have improved the outbreak surveillance and detection process, there has been a peak in the number of recalls related to foodborne and environmental illnesses within the last several years. With growing research, public health investigators have begun looking into climate impact as the tense yet prudent conversation surrounding the impacts of climate change continues to rage on. The connection between environmental circumstances and future food-borne outbreaks is a convoluted relationship that has limited but valuable research; therefore, we hope to open the door to further uncovering the correlation between the two to help educate future decisions and understandings of agricultural practices and climate changes.

Completing the analysis will be a multi-step process that first involves gathering and preparing the raw data from various sources within the University of Arizona research archives. We will also be looking at data from sources such as the CDC NORs Dashboard and National Weather Service to gather recall and environmental data. The project will also encapsulate data cleaning and analysis through powerful tools, such as Python, R, and SQL, and visualization through R and a Tableau Dashboard. Various statistical methods such as regression analysis and clustering and supervised and unsupervised machine learning will be applied to categorical and continuous factors to reveal any patterns or notable relationships described and visualized in the final statistical report.

Team Member	Feature responsibility
Alex Alonso	*Data collection and Tableau dashboard and Final Report*
Janessa Davis	*Machine Learning and Analysis and Final Report*
Nicholas Filiberti	*Data collection and Tableau dashboard and Final report*
Khusniddin Abdullaev	*Data cleaning and Processing and Final Report*
Xiunan Song	*Python for Data cleaning and Processing and Final Report*

Table 1 Preliminary Subsystem Responsibilities

2. Literature Review/Market Research

The Literature Review was written by Janessa Davis.

Food safety and climate change have continuously been key topics of scientific research and public concern as the agricultural industry continues to expand and daily temperatures continue to rise. With health concerns continuing to peak, scientific investigation has begun to uncover relationships between climate change and related food- and water-borne pathogens. As investigating the role of environmental elements has been a research topic for several agricultural scientists and investigators, notable links and patterns have been identified between environmental conditions and common food-borne illnesses such as *Salmonella, Listeria*, and non cholera *Vibrio* (Lake, 2019[1]). With the prediction of higher temperatures, heat waves, and storm surges, creating connections between climate conditions and outbreak patterns has improved with the introduction of surveillance tools such as the CDC NORs dashboard and quantitative microbial risk assessment (QRMA[2]) (Schijven, 2013[3]).

Preliminary microorganism detection through processes such as the QRMA has allowed for associations such as increased precipitation and rises in *Salmonella* to be discovered (Schijven, 2013). Thus, research begins to illuminate the seasonality and weather trends between pathogens and environmental circumstances to have greater predictive power in the public health sphere. A greater understanding of the role of environmental and weather factors increases the recognition and priority of food safety in the agricultural production industry (Semenza, 2011[4]). Through both contextual assessments and quantitative analysis, notable patterns have been identified between ecological circumstances and prevalent food- and water-borne pathogens in northern European regions (Semenza, 2011). Therefore, seasonality predictions have become more conceivable with pattern recognition. Weather patterns such as an increase in precipitation or a drought period due to increased temperatures can be telltale indicators for preparedness techniques that should be considered to curb future outbreaks and recalls (Lake, 2019).

Although the conversation about food safety and climate change continues to ramp up within their respective industries and fields, research and experimentation surrounding the connection between the two still offer predictive potential for future public health concerns (Schijven, 2013). Therefore, this research project will provide a new perspective by utilizing the CDC recall and outbreak data to investigate the entire farm-to-table process and those things that can disrupt and contaminate it.

---

[1] Lake, I. R., & Barker, G. C. (2018). Climate Change, Foodborne Pathogens and Illness in Higher-Income Countries. Current environmental health reports, 5(1), 187–196. https://doi.org/10.1007/s40572-018-0189-9

[2] European Centre for Disease Prevention and Control (Ed.). (2016, June). ECDC quantitative microbial risk assessment (QMRA) tool. Climate Adapt. Retrieved March 10, 2024, from https://climate-adapt.eea.europa.eu/en/metadata/tools/ecdc-quantitative-microbial-risk-assessment-qmra-tool

[3] Schijven, J., Bouwknegt, M., de Roda Husman, A.M., Rutjes, S., Sudre, B., Suk, J.E. and Semenza, J.C. (2013), A Decision Support Tool to Compare Waterborne and Foodborne Infection and/or Illness Risks Associated with Climate Change. Risk Analysis, 33: 2154-2167. https://doi.org/10.1111/risa.12077

[4] Semenza, J. C., Herbst, S., Rechenburg, A., Suk, J. E., Höser, C., Schreiber, C., & Kistemann, T. (2012). Climate Change Impact Assessment of Food- and Waterborne Diseases. Critical reviews in environmental science and technology, 42(8), 857–890. https://doi.org/10.1080/10643389.2010.534706

While research has been conducted regarding European regions and seasonality, connections here in the United States have not been understood to the same degree. Therefore, this research seeks to acknowledge and discover patterns in the US that can help illuminate outbreak increases seen throughout the years. This data analysis can offer further insight into climate changes witnessed in major US production regions and pathogen spikes to improve public health and food safety and awareness for both agricultural producers and consumers alike.

3. Research Project Deliverables

The Research Project Deliverables were written by Alexander Alonso and edited by Janessa Davis.

Research Deliverables

- Research paper that…
    - Summarizes the data, why it was chosen, the analytics methods, the final results and visualizations from the analysis and predictive model, and the final findings
    - Will have 5 peer-reviewed citations with methods, results, and discussion
    - Will have data collection and cleaning done through Python and SQL
    - Has analysis is done through Python and R
- Analysis and Pattern Recognition
    - Using unsupervised machine learning to avoid placing assumptions on data
        - K-means clustering - determining outbreak, environmental, and seasonality patterns within major regions (R/Rstudio)
        - Decision Tree/Random Forests - predicting outbreak-based seasonality and time-based factors, weather and climate conditions, and region (R/Rstudio)
    - Outbreak and pathogen data will be pulled from the CDC NORS dashboard
    - Weather and climate data pulled from National Weather Service and related sites using scrapping and APIs along with PRISM R-package and dataset
- Prediction Model for Future Food Outbreaks
    - Classification model (traditional/penalized logistic regression and/or linear discriminant analysis (LDA)) using relevant factors to predict the likelihood of a food-related outbreak
        - wanting a model that has at least 60% accuracy (R^2 value of 0.6 or higher)
        - Will use traditional/penalized logistic regression or linear/quadratic discriminant analysis (LDA/QDA) depending on the distribution and nature of the data
    - Done through R or Python depending on which platform better caters to the data
    - In the case that the data provides a null or unsuccessful model, additional visualizations will be created and statistical analysis and summary will be completed
- Tableau Dashboard
    - Interactive dashboard that allows viewers to see seasonality patterns and summaries in major regions, outbreak patterns, and the connections found between outbreaks and climate

4. Project Timeline & Gannt Chart

The Executive Summary was written by Nicholas Filiberti and Janessa Davis.

Milestone	Date
Team Formation	1/29/24
Signed proposal	3/13/2024
Data Collection	3/14/2024
Data Upload	3/15/2024
Initial Data Investigation	3/18/2024
Data Cleaning and Preprocessing	3/22/2024
Machine Learning Analysis	4/5/2024
Prediction Model Creation	4/5/2024
Tableau Data Upload	4/8/2024
First Draft of Report	4/15/2024
Tableau Skeleton	4/17/2024
Final Tableau Dashboard	4/22/2024
Poster Demo	4/23/24
Final Draft of Report	4/24/2024
iShowcase	5/1/24

Table 3: Milestone Schedule

Gantt Chart: ISTA498_Capstone_GanntChart

5. Ethics

The Executive Summary was written by Nicholas Filiberti and Xiunan Song.

#	Question	Generally	Data Breach
1	Could a user sell drugs or other illegal items on your platform?	No	No
2	Could a user of your platform engage in sex trafficking?	No	No
3	Could a user sell class notes or cheat on their homework on your platform?	No	No
4	Could a stalker use your project to find someone?	No	No
5	Could your app be used to spy on or track individuals?	No	No
6	Could your app/software access the camera or microphone and record things without users being aware?	No	No
7	If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way?	No	No
8	Could your algorithm promote material that would traumatize or upset individuals?	No	No
9	Would your users be upset if the data you collect was given to someone else?	No	No
10	Could a data leak potentially lead to identity theft?	No	No
11	If your site was hacked, would users of that product potentially lose their job, spouse, or family?	No	No
12	Should there be an age limitation on your product?	No	No
13	Could someone use your product to find, contact, and potentially commit elder abuse?	No	No
14	If the data on your platform was breached, could it be used to blackmail the users?	No	No

15	Does the existance of your project imply that a particular racial group, gender, religion or other protected category is inherently bad, gross, or unwanted?	No	No
16	Could your product be used to commit hate crimes against a specific group?	No	No
17	Does the primary content of your game or algorithm focus on something considered deeply unethical?	No	No
18	Does your game or software contain race, gender, or other stereotypes?	No	No
19	Could users of your app scam other individuals?	No	No
20	Is your particular algorithm biased towards predicting correctly only for one race, gender, or other group?	No	No
21	Are the users of your project, players of your game, or those being surveyed for your data aware of how their data will be used?	No	No
22	What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it?	No	No
23	Does the use or purchase of your data potentially contribute to a dangerous group or regime?	No	No
24	Could your virtual reality environment cause injury to the user?	No	No
25	Are your study participants or game players aware that their data will be collected and used?	No	No
26	Does your game or app contain addictive design elements without benefit to the user?	No	No
27	Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment?	No	No

28	Could your research outcomes harm an individual or entity?	No	No

6. Approvals

The signatures of the people below indicate an understanding of the purpose and content of this document by those signing it. By signing this document, you indicate that you approve of the proposed project outlined in this Statement of Work, the division of work, and the Ground Rules, and that the next steps may be taken to create a Product Specification and proceed with the project.

Approver Name	Title	Signature	Date
Janessa Davis	Team Project Manager		
Alex Alonso	Team Member		
Nicholas Filiberti	Team Member		
Khusniddin Abdullaev	Team Member		
Xiunan Song	Team Member	Xiunan Song	3/13/2024
Dr. Kristen Pogreba Brown	Advisor		
Dr. Erika Austhof	Advisor		
Prof. Michael McKisson	Instructor		

Section	Author	Word Count
Executive Summary	*Janessa Davis, Khusniddin Abdullaev, Xiunan Song*	*293*
Literature Review	*Janessa Davis*	*434*
Research Project Deliverables	*Alex Alonso*	*270*
Project Timeline & Gannt Chart	*Nicholas Filiberti, Janessa Davis*	
Ethics	*Nicholas Filiberti, Xiunan Song*	

7. Appendix

A. Advisor Engagement

1) Project Team Responsibilities

- The Project Manager will set up and facilitate a weekly call/meeting with the Faculty Advisor. The Project Team will provide weekly status updates to the Faculty Advisor including upcoming deliverables, critical issues, and any adjustments to the Project Plan.
- Documents will be provided to the Faculty Advisor with adequate time for review and signature. The time necessary for review will be agreed with the Advisor. The minimum review time will be 3 days prior to the document due date.
- Design files will be provided to the Faculty Advisor as requested in a format agreed to with the Advisor.
- Support requirements will be clearly requested from the Faculty Advisor with the dates required and an adequate time for fulfilling the request.
- Modifications requests to the Project Plan by Faculty Advisor will be reviewed and agreed to within 1 week of the request.

2) Faculty Advisor Responsibilities

- The Faculty Advisor will provide knowledge and expertise to help the group stretch their skills.
- The Faculty Advisor will participate in a weekly or bi-weekly call/meeting with the Project Team to review the project status, upcoming deliverables, priorities, issues, and progress to the agreed Project Plan.
- The Faculty Advisor will provide document review, feedback and approval, rejection, approval with contingencies with adequate time for the Project Team to meet the course due dates.
- The Faculty Advisor will provide feedback to requested support requirements from the Project Team. This includes feedback and guidance on design implementations decisions, design files, test plans, test procedures and test results.
- The Faculty Advisor shall provide technical advice and guidance to the Project Team answering inquiries approximately 1 hour per week.
- Modifications to the Project Plan by the Project Team will be resolved and documented within 1 week of the request.
- Grade the finalized project using a skill-based rubric
- Attend iShowcase in May.

B. Ground Rules

As a team and as individual team members, we agree to:

1. **Stay focused on our objectives and goals.**
   Each time the team meets, we will clearly define our objectives and desired outcomes at the beginning of the meeting. We will politely remind team members if we are getting off track.

2. **“Sidebar” any issues that are relevant but not consistent with the immediate objectives.**
   Occasionally, important matters are raised that are not relevant to the immediate goals of the meeting. To keep the group on track, but avoid losing the issue, create a “sidebar” where these topics can be listed and discussed later.
3. **Listen when others are speaking.**
   We will listen and consider others’ input before adding our own comments.
4. **All viewpoints will have an opportunity to be heard.**
   We understand that some team members may be quieter than others. We will make an effort to get each team member’s viewpoint and that no one dominates the discussion.
5. **Differences of opinion will be discussed respectfully**
   We will identify areas of agreement before assessing areas of disagreement. We will encourage each other to look beyond our own point of view. We will discuss different ideas respectfully. As a team, we will weigh the merits of different opinions and agree on a process for choosing a direction. All team members will respect and follow the decision or direction.
6. **Look for the good points in new ideas.**
   We will endeavor to explore the value in each idea as we assess and select our path forward.
7. **Focus on the future, not the past.**
   We will use our past experience to inform our decisions, but focus the discussion on the future objectives. Blame for past performance is counterproductive, we will focus on finding solutions.
8. **Agree upon specific action items and next steps.**
   At the end of each meeting and discussion, we will summarize and agree on specific next steps, action items and assignments.
9. **Accountability**
   As team members, we will each be responsible for our individual assignments and contribution to achieving the team objectives and goals. We will honor our responsibilities and not let our team members down.