  **DAEN-690 PROJECT REPORT**

**FOOD TRACEABILITY STUDY WITH GRAPH CONVOLUTIONAL NETWORKS**

**Team Members:**

**Aidin Ziaee - Scrum master**

**Sindhuja Nainapally - Product owner**

**Arun Kumar Gantyada - Github owner**

**Jashwant Naidu Lakkakula - Developer**

**Vaishnavi Kadapathri - Developer**

**Contents**

**Table of Contents**

**ABSTRACT**

**INTRODUCTION**

**SECTION 1: PROBLEM DEFINITION**

**1.1** **Background**

**1.2** **Problem Space**

**1.3** **Research**

**1.4 Solution Space**

**1.5 Problem Objectives**

**1.6 Primary User Stories**

**1.7 Product Vision**

**SECTION 2: DATASETS**

**2.1** **Overview**

**2.2** **Field Description**

**2.3** **Data Context**

**APPENDIX A: GLOSSARY**

**APPENDIX B: GITHUB REPOSITORY**

**Overview**

**GitHub Repository Link**

**GitHub Repository Contents**

**APPENDIX C: RISKS**

**Sprint 1 Risks**

**Sprint 2 Risks**

**APPENDIX D: AGILE DEVELOPMENT**

**Scrum Methodology**

**Sprint 1 Analysis**

**ABSTRACT:**

The U.S. Centres for Disease Control estimate that there are 48 million instances of foodborne infections each year, which represents a huge worldwide health concern. Intricate supply networks make it difficult and time-consuming to identify the source of food poisoning. The goal of this project is to resolve this problem by creating a Product Tracking System that replicates data and automates the detection of probable contamination sources inside the food supply chain network.

The primary issue raised by this study is the absence of effective and precise traceability across the food supply chain, which makes it difficult to quickly identify and contain outbreaks of foodborne illness. By utilizing the Food Traceability Rule of the FDA, the EPCIS Data Standard, and the Core Business Vocabulary (CBV), the research closes this gap. A realistic model of the food supply chain is developed using data simulation techniques and machine learning algorithms, which also identify possible sources of contamination.

This study has important ramifications for enhancing food security, safeguarding public health, and avoiding monetary losses brought on by foodborne illness outbreaks. The suggested method may completely alter how foodborne illness outbreaks are monitored and controlled if it is put into practise on a broader scale.

Food items are tracked and documented as they move through the supply chain from manufacturing to consumption, which is a crucial component of the food business. Strong traceability systems are essential at a time when commerce is more globalized and consumer worries about food safety and authenticity are at an all-time high. In order to improve traceability practices, this study examines the state of food traceability today, finds problems, and suggests creative solutions. The study's objective is to develop food traceability procedures and promote the more general objectives of food safety and supply chain effectiveness by fusing rigorous research with real-world application.

This project recognizes the critical importance of food traceability in ensuring the safety and authenticity of the food supply. It also acknowledges the potential of emerging technologies, like GCNs, to revolutionize the way we approach traceability in complex supply chain networks. By combining rigorous research and practical implementation, our study aims to contribute to the advancement of food traceability practices and support the broader goals of food safety and supply chain efficiency.

**REPORT**

**Section 1 : problem definition**

**1.1 Background**

Food traceability is a critical component of our modern food supply chain, serving as a crucial tool in ensuring the safety, quality, and transparency of the food we consume. In an era where consumers are increasingly conscious of the origin and journey of their food, traceability has emerged as a fundamental practice for both producers and consumers. This multifaceted concept involves tracking and documenting the movement of food products from their point of origin through every step of the supply chain until they reach the hands of consumers. Whether it's a juicy apple, a tender cut of meat, or a bag of rice, consumers want to know where their food comes from and trust that it meets stringent safety standards. In this era of globalization, food traceability has become more complex and challenging, yet it has also become more essential than ever before.

The importance of food traceability cannot be overstated, as it plays a pivotal role in safeguarding public health. In an interconnected world where food products cross borders with ease, the potential for contamination and foodborne illnesses is a constant concern. Food traceability systems enable rapid identification of the source of contamination, helping authorities to swiftly recall affected products and prevent widespread outbreaks. Furthermore, traceability allows producers to implement targeted quality control measures, ensuring that the food products they deliver to consumers meet the highest standards of safety and quality. By providing a transparent and accountable system, food traceability fosters consumer confidence, assuring them that the food they purchase is wholesome and free from harmful substances.

The ingestion of food or drink contaminated with dangerous germs, such as bacteria, viruses, or parasites, or chemicals, such toxins or metals, can cause foodborne illnesses, a serious worldwide health problem. With symptoms ranging from slight stomach pain to disorders that pose a serious risk of death, these illnesses can be mild or severe. According to the Centres for Disease Control and Prevention (CDC), there are 48 million instances of foodborne illnesses each year in the United States alone, which result in 128,000 hospitalisations and 3,000 fatalities. [1] These concerning numbers highlight the urgent requirement for efficient food tracking systems. In order to protect public health and lessen the financial burden associated with these illnesses, such systems can assist in locating the origins of epidemics, ascertaining their causes, and preventing similar future occurrences.

Public health and the economy may be greatly impacted by comprehending these chains and putting prevention measures in place. A thorough understanding of the food supply chain can aid in locating possible locations of contamination, enhancing food safety procedures and disease tracking systems. Additionally, it can support the creation of laws and policies that safeguard the public health by preventing foodborne infections. Effective food supply chain management may lower the financial consequences of foodborne disease outbreaks, including medical expenses, lost productivity, and trade disruptions. Therefore, in terms of preventing foodborne diseases, the significance of comprehending and controlling food supply chains properly cannot be emphasized.

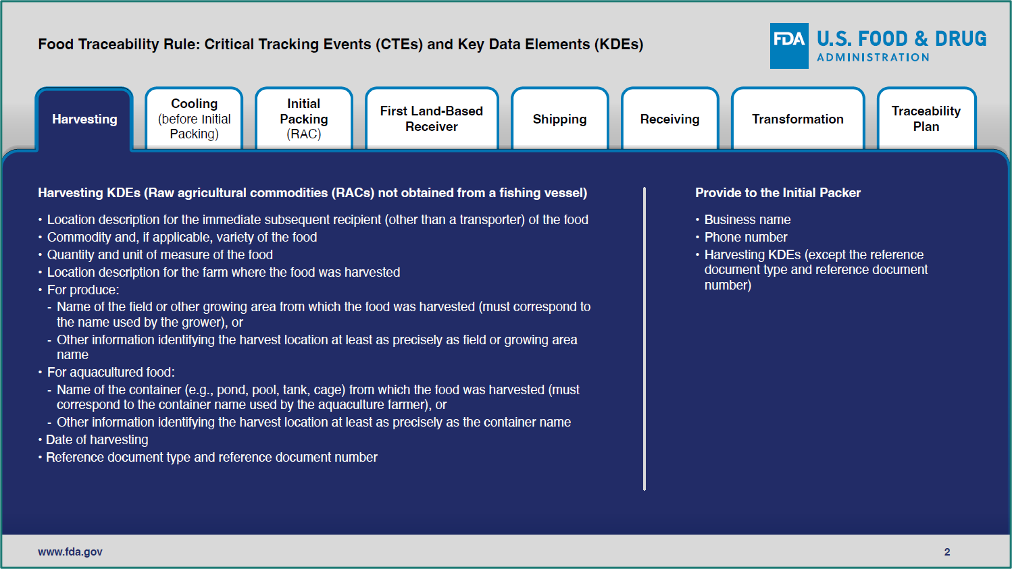
In addition to being desirable qualities, the food supply chain's openness and traceability are crucial elements in guaranteeing food safety and public health. A clear, unobstructed picture of each step in the manufacture and transportation of food is made possible by the food supply chain's transparency. From the sourcing of raw materials through the last step of delivering the product to the customer, it enables close examination of each stage. By ensuring that safety standards and best practices are followed across the whole food supply chain, this visibility helps to reduce the risk of contracting foodborne diseases.

Contrarily, traceability refers to the capacity to track and trace a food product's movement across the supply chain. In the case of food contamination, this skill is very important. Effective traceability enables quick and efficient food recalls by making it feasible to identify the source of contamination. This not only stops the epidemic from spreading further, but it also helps pinpoint the underlying source of the contamination, enabling remedial actions to stop similar incidents from happening in the future.

The Food Traceability Rule was developed by the Food and Drug Administration (FDA) as part of the Food Safety Modernization Act (FSMA) to improve supply chain transparency and traceability. For individuals that produce, prepare, pack, or hold items included on the Food Traceability List (FTL), this rule demands extra record-keeping obligations in addition to those already mandated by law.

This rule's core requirement is that these entities keep records with Key Data Elements (KDEs) connected to particular Critical Tracking Events (CTEs). These KDEs and CTEs act as crucial signposts in the food supply chain, identifying significant moments and provide crucial information that may be utilized to track the path of a food product.

Take the crucial occurrence of harvesting as an example. The accompanying KDEs are depicted in Figure 1 and include the location of the following food receiver, the amount and unit of measurement of the food, the location of the farm or facility where the food was produced, and the date of production. When data is reliably collected and immediately provided, supply chain transparency is improved. In the case of a foodborne epidemic, it enables quicker identification of the impacted nodes in a supply chain, assisting in the prevention and control of such outbreaks.



Through its focus on openness and traceability, the FDA's Food Traceability Rule offers a strong foundation for controlling food safety hazards throughout the supply chain. It emphasises how crucial prompt reporting and proper record-keeping are for avoiding foodborne diseases and safeguarding the public's health.

Enhancing the efficiency of food monitoring systems increasingly depends on the application of cutting-edge technology and data science methodologies. Artificial intelligence's subset of machine learning algorithms has been quite effective in this area. These algorithms can examine large, complicated information to find patterns and connections that may be difficult for human analysts to see.

These patterns can offer insightful information about possible weak points in the supply chain, enabling the identification of potential sources of contamination. As it enables the deployment of preemptive measures aimed at preventing foodborne diseases, this predictive skill is of enormous significance. It changes the paradigm from a proactive one, where possible problems are discovered and treated before they may cause harm, to a reactionary one, where measures are made after an incident has occurred.

Furthermore, it is impossible to stress the importance of data visualisation tools in this situation. These techniques might transform intricate data patterns into simpler and clearer visual representations. These technologies make information visually appealing so that stakeholders may easily understand complex ideas or spot novel trends. Data visualisation can make the narrative behind the data come to life, from heat maps that give a visual depiction of the data to scatter plots that may show relationships between different variables. As a result, it becomes a crucial instrument in the management of food safety, assisting in the decision-making processes and improving the general comprehension of supply chain dynamics.

Another important development in the age of digital transformation is the integration of these systems with cloud-based platforms. Real-time data exchange and cooperation amongst various food supply chain stakeholders are made possible by cloud-based technologies. This comprises, among others, farmers, processors, merchants, and regulators. By considerably accelerating the reaction time in the case of an epidemic, this capacity can improve the system's ability to stop the spread of foodborne diseases.

In conclusion, it is a challenging but crucial responsibility to comprehend and manage food supply networks. It necessitates a multifaceted strategy that includes adherence to legal requirements such as the FDA's Food Traceability Rule, the adoption of reliable monitoring systems, and the use of cutting-edge data science methodologies. When combined properly, these components may provide a thorough food safety management system that not only assures the food's safety and quality but also helps to safeguard the general public's health.

We are more prepared than ever to meet the ongoing challenge of preventing foodborne diseases because of technology and data science. To improve food safety, safeguard public health, and create a more resilient and sustainable food system going ahead, it is critical that we keep coming up with new ideas and using these technologies.

Beyond its crucial role in food safety, traceability also contributes to sustainability and environmental responsibility. As global populations continue to grow, so does the demand for food. This puts immense pressure on agricultural and food production systems, leading to concerns about overexploitation of resources and environmental degradation. Food traceability helps address these concerns by enabling better management of supply chains, reducing food waste, and promoting sustainable practices. When we can trace the journey of a food product, we can make informed decisions about its environmental impact, considering factors such as transportation, production methods, and packaging. This knowledge empowers consumers to choose products that align with their values and environmental priorities.

Furthermore, food traceability has a profound impact on the economic aspects of the food industry. Producers who implement robust traceability systems often gain a competitive edge by differentiating their products in the market. Consumers are increasingly seeking out foods with clear and verifiable origins, and this demand can translate into higher sales and brand loyalty for companies that can provide this information. Additionally, traceability can streamline supply chain operations, reducing costs associated with recalls, compliance issues, and inefficiencies. It also opens up new opportunities for innovation and collaboration within the industry, as data generated through traceability can be harnessed to optimize production, distribution, and marketing strategies.

In summary, food traceability is a multifaceted concept that impacts various aspects of the food industry, from safety and sustainability to consumer choice and economic competitiveness. In an era characterized by a globalized food supply chain and increasing consumer awareness, the need for robust traceability systems has never been greater. By ensuring the safety, quality, and transparency of our food, traceability plays a vital role in nourishing a growing world population while safeguarding public health and the environment.

**1.2 Problem Space:**

Our project's underlying subject is food safety, with an emphasis on improving the supply chain's transparency and traceability. While traceability enables the identification and monitoring of a food product's movement across the supply chain, transparency offers a clear picture of each activity involved in food manufacturing and delivery. This feature is especially important in the event of food contamination since it makes it possible to identify the source of the contamination, facilitates efficient food recalls, and stops the spread of the epidemic.

However, guaranteeing food safety is a difficult undertaking with many difficulties. The supply chain's incompleteness is one of the biggest problems.

The food industry faces challenges in establishing and maintaining effective food traceability systems that can reliably and efficiently track the provenance, handling and distribution of food products. This problem statement seeks to address the overarching issues within food traceability analysis, including the limitations of existing systems, technological barriers, regulatory compliance, and the impact of traceability on various stakeholders.

The present food tracking systems still have large gaps despite these steps. The absence of sufficient real-world data owing to security concerns and a lack of rules is one of the difficulties in supply-chain analytics.

The precision of synthetic data points, which are created to make up for this deficiency, is frequently lacking. When training and assessing data models, using such intentionally created points might result in subpar model performance.

The tracking of perishable commodities that are in the supply chain is another issue that is unique to the food business. From harvesting to procurement to their final destination in retail establishments or restaurants, perishable foods can get contaminated at any stage in the supply chain. The FDA officials must manually examine the supply chain data in the event of food contamination in order to pinpoint the origins and utilize the information to check such places. To find any source or links that may lead to contamination, however, requires sifting through the whole supply-chain network, which is quite difficult sometimes.

By creating a strong product tracking system (PTS) that makes use of cutting-edge technology and data science methodologies, our project intends to address these issues. The system will be able to do two things in particular:

1. Track-back capability: The system must be able to track backwards from a point of contamination to identify the original sources. This is crucial for pinpointing the root causes of an outbreak.

2. Forward tracking: The system should also be capable of tracing forwards from a known contamination source to identify all potentially affected nodes downstream. This enables preventative actions to contain the spread of pathogens.

We must make use of cutting-edge technology and approaches to create this sophisticated supply chain analytics system:

* Gather and examine data from actual supply chains to comprehend node interactions. Machine learning models will be trained using this data to derive supply chain mappings.
* Use synthetic data if real-world supply chain information cannot be obtained. Our product adheres to EPCIS (EPC Information Services) requirements and maintains realism by using synthetic data that was created using a Python script.
* In order to provide visibility between organizations and throughout a supply chain, GS1 developed the EPCIS data sharing standard. in order to enable the recording and exchange of interoperable information regarding status, location, and traceability throughout the supply chain, it aids in providing the "what, when, where, why, and how" of products and other assets.
* To model the linkages in the supply network, conduct research and put graph databases and graph analytics tools to use. The fast traversal required for forward and backward tracking is made possible by graph databases.
* When real-world data is lacking, experiment with simulation models to create synthetic but realistic supply chain data for testing. Validation must be done carefully to guarantee usefulness.
* Create user-friendly user interfaces, interactive dashboards, and data visualizations to showcase supply chain analytics insights and encourage data exploration by stakeholders.
* Implement robust data security protections to maintain confidentiality of sensitive supply chain data while enabling analysis.

Successful implementation of this system will provide improved supply chain visibility and traceability to augment food safety. The ability to rapidly pinpoint sources and pathways of contamination will lead to faster outbreak containment and prevention. This will help reduce the substantial health and economic burdens of foodborne diseases.

## **1.3** **Research**

We combed through a plethora of publications and sources dispersed over the broad expanse of the internet in our effort to investigate the dynamic landscape of food safety. We set out with a single goal in mind: to find a significant connection between the problems that now plague the field of food safety and the creative solutions that these academic publications and sources may be able to provide.

The risks of contracting a foodborne illness have increased due to the world's food supply chains' ever-increasing complexity, impacting people, businesses, and entire countries. As a result, our research intended to close the knowledge gap that exists between the challenges that people encounter every day and the vast body of knowledge that experts, scientists, and researchers have accumulated.Each document, source, and study served as a component of the full picture of the terrain that we painstakingly put together.

We will describe our findings in the parts that follow, illuminating the complex relationships between the current issues with food safety and the prospective remedies that these academic works can provide. Every paper we read helped us understand concepts from a different angle, enabling us to tackle global concerns with food safety in a more educated and proactive manner.

Research Article 1: "Smart, Digitized Food Supply Chain through Self-Learning and Self-Adaptive System”

Summary of the Paper:

The paper discusses a concept to reduce food waste and improve food safety in the food supply chain using adaptive systems. It focuses on smart sensors that monitor food conditions and machine learning for predicting food quality and shelf life. The main objective is to integrate these technologies with blockchain for real-time data analysis, allowing proactive food supply management.

Connection to Existing Problem:

* The paper addresses the challenges in food safety and traceability by proposing a system that combines smart sensors, blockchain, and real-time data analysis.
* It aligns with the problem of incomplete supply chain information, gaps in tracking systems, and the need for rapid identification of contamination sources.
* Using blockchain for transparent data sharing and integrating sensors in packaging helps prevent the sale of spoiled food.
* The emphasis on real-time tracking and shelf life predictions aligns with the goal of preventing foodborne diseases and improving food safety.

Research Article 2: “Study on Microwave Food Contamination Examination Based on Machine Learning Approach”

Summary of the Paper:

This paper explores the application of machine learning in examining microwave food contamination. It uses the Gauss-Newton method to assess the information in microwave signals. Then, it employs a Support Vector Machine (SVM) for food contamination determination and a Back Propagation (BP) neural network to detect foreign matter's position and size. The study demonstrates that microwave signals can effectively be used for non-invasive food contamination detection, and machine learning methods like SVM and BP neural networks are suitable for this purpose.

Connection to Existing Problem:

* The paper addresses the issue of food contamination, which poses a significant challenge to food manufacturers in terms of quality assurance, brand reputation, and financial costs associated with recalls.
* It explores emerging imaging techniques, including microwave imaging, as a means to ensure food quality and safety.
* The use of machine learning tools like SVM and BP neural networks aligns with the need for efficient and accurate food contamination detection, improving food safety.
* Microwave imaging's non-invasive nature makes it a valuable tool for detecting food contamination, aligning with the overarching problem of enhancing food safety and quality assurance.
* The paper highlights the potential of microwave technology and machine learning to mitigate food safety incidents and improve the overall safety of the food supply chain.

Research Article 3: “Research on the Application of Big Data in Smart Food Safety”

Summary of the Paper:

This paper focuses on the application of big data in smart food safety. It begins by examining the sources and characteristics of big data in the food industry. It discusses data from the food supply chain, production processes, logistics, sales, and more. The paper emphasizes that big data in the food industry exhibits the 4V characteristics of volume, variety, velocity, and value density.

Connection to Existing Problem:

* Food safety is a critical global concern that threatens public health. The paper addresses this issue by proposing the use of big data and information technology to enhance food safety.
* The paper highlights the need for a unified food safety computing service platform to provide comprehensive, accurate, and authoritative food safety information to various stakeholders.
* It emphasizes the importance of analyzing and predicting food safety, offering information services to consumers, and supporting government supervision and decision-making.
* The sources of big data in the food industry, such as the supply chain, production processes, and regulatory standards, align with the need for comprehensive data to ensure food safety.
* The paper discusses data preprocessing, cleaning, integration, and reduction, which are essential steps to make big data suitable for analysis and decision-making in food safety.
* Parallel big data mining and analysis are introduced as crucial techniques for efficiently processing and extracting insights from large-scale food safety data.
* The paper presents a basic framework for big data processing in smart food safety, including data collection, processing, analysis, and user application layers.
* It outlines application contexts for big data in food safety, such as food information electronic traceability systems, which help track and manage food supply chains for enhanced safety and quality assurance.

## **1.4** **Solution Space**

The solution space encompasses addressing two primary challenges. Initially, it involves generating simulated supply chain data due to the absence of actual real-world supply chain data.Exploring various data population techniques to generate realistic data points that mimic real-world supply chain scenarios.Utilizing pre-existing data generators to create additional data from the limited available real-world data. Gaining insights from sample real-world data to understand input data characteristics.

Approaches to Data Generation:

* Python Workflow with 'Faker' and 'Random' Packages
* User-Friendly Interface with 'Dash' Library
* Python API with Customizable Features

Implementing a terminal application that accepts user input for the number of business entities and food items to simulate.Offered multiple approaches to data generation, providing flexibility and customization.Ensured data completeness, uniqueness, and conformity to industry standards.Facilitated ease of data generation through user-friendly interfaces.Enabled future use of the Python API for similar projects.

Subsequently, it centers around developing robust models utilizing the generated data to predict potentially contaminated points within the supply chain, with the ultimate goal of preventing future outbreaks.Choosing a graph-based storage system to capture and maintain the interconnected relationships within the supply chain data, ensuring data integrity and flexibility for analysis.Opt for Neo4j, a widely-used graph database, to implement the solution. Neo4j Desktop is selected as it provides a user-friendly interface and essential tools for graph-based data management and analysis.Initially considering hosting the database in a cloud instance but decide to use Neo4j Desktop for local deployment, which bundles Neo4j Browser, Neo4j Bloom, and graph data science libraries.To Develop a Python workflow to load the supply chain data into the Neo4j database. This workflow is responsible for transferring data generated using 'faker' and 'random' packages into the database.Setting up a local Neo4j database with appropriate authentication (username and password) in Neo4j Desktop.Executing the Python workflow to ingest generated data in CSV format into the Neo4j database. Ensure that data integrity and structure are maintained during this process.Utilize Cypher queries within Neo4j Browser to perform various data analysis tasks. Cypher is a query language designed for graph databases, making it suitable for extracting insights from the supply chain graph.Visualizing the supply chain network using Neo4j Bloom, a graph visualization tool that enables a clear representation of interconnected data.Leverage the graph structure of the data stored in the Neo4j database to build machine learning models and Utilize Cypher queries to extract relevant features from the graph for logistic regression and XGBoost models.

## **1.5** **Project Objectives**

* To identify where in the supply chain the contaminants are likely to occur.
* To forecast potential model outcomes using node embedding methods and without them.
* To utilize the application of ML models to forecast the business value of a foodborne outbreak at particular supply chain nodes.
* To build a database of graphs to help with traceback and trace forward diagrams. The team should be able to learn about graph database technology
* To create an interactive, visual diagram that represents the full supply chain for certain products and food items utilizing graph database technology.
* The entire solution entails storing the supply chain data in a graph database like Neo4j, using this data to create and put into use machine learning models that accurately forecast polluted nodes in a supply chain, and then integrating all of these components.

**1.6 Primary User Stories**

* Applying ML models including exploring more on Graph Convolutional Neural Network.
  + GCNs are a type of neural network designed for processing data represented as graphs. In a graph, nodes represent entities, and edges represent relationships between those entities. GCNs are particularly useful when dealing with data that has a relational or network structure, such as social networks, recommendation systems, biology, and more. We would like to take advantage of its application to be able to promptly track the contaminated nodes and prevent the food-outbreak and optimally predic the next contaminated nodes.
* Exploring Node Embedding algorithms such as building on the HashGNN with Neo4j, DeepWalk, FastRP, GraphSAGE using the Skeptral library in Python
  + Node embedding algorithms are a class of techniques used to represent nodes in a graph as continuous, low-dimensional vectors. These embeddings capture structural and relational information about the nodes in a way that is suitable for machine learning tasks. Node embeddings have applications in recommendation systems, link prediction, community detection, and more. We are aiming to take advantage of these algorithms including HashGNN, DeepWalk, FastRP, and GraphSAGE with benefits of Skeptral library.
* Exploring Oversampling techniques or adjusting the contamination rate to increase the chances of contamination.
  + Anomaly detection is a data analysis technique that focuses on identifying patterns in data that do not conform to expected behavior. These unusual patterns are often referred to as anomalies or outliers. Outliers can be indicative of errors, fraud, or other unusual events in the data. The goal of anomaly detection is to distinguish between normal and abnormal data points. Exploring oversampling techniques and adjusting the contamination rate are strategies used in anomaly detection to address imbalanced datasets and increase the chances of detecting anomalies. Oversampling creates synthetic anomalies, balancing the dataset, and potentially improving model performance which we are aiming to apply in our models.
* Being able to track back to the origin(s) of the outbreak.
  + Given a complete network, and known node(s) that food outbreak happens (for example, a grocery store) , If a food is made by multiple ingredients, and each ingredient was produced by different farm, then prioritize the risk level for each farm so that inspector could go to highest risk farm first to conduct inspection. This also would be an application of Graph Network Analysis. We need to study similar behavior network connectivity.
* Being able to predict the missing edges given an incomplete network.
  + Scenario: If we already have a complete food supply chain network and we know an node that has a food outbreak, it is relatively easier to use trace-back algorithm to find out the origin (farm, food manufacturer, so on) of the contaminated food ingredient. But in real word, it can be very hard to construct a complete network, it is likely that some connections (edges) between nodes are missing because suppliers miss to report some of the shipments. In this case, the task is to predict those missing connections. We should be Comparing the similarity between the subsets of the networks and understand how we can apply them

## **1.7** **Product Vision**

**Transforming Food Traceability for a Safer, Sustainable, and Transparent Food Supply Chain.**

A project Vision for food traceability would encompass the overarching goals , objectives, and principles that guide the development and implementation of a comprehensive food traceability system. Food traceability is essential for ensuring food safety, quality, and sustainability, as well as for meeting consumer demand for transparency in the food supply chain. Here’s the project vision for food traceability.

**1. Ensuring Food Safety:**

* Our project envisions a world where every consumer can trust that the food they eat is safe and free from contamination. We aim to achieve this by implementing a robust food traceability system that allows for rapid identification and recall of unsafe products.

**2. Enhancing Food Quality:**

* We strive to improve food quality and freshness by enabling real-time monitoring and tracking of food products throughout the supply chain. This ensures that consumers have access to the highest quality and freshest products.

**3. Promoting Sustainability:**

* Our vision includes a commitment to sustainability. We aim to reduce food waste, promote sustainable farming practices, and minimize the environmental impact of the food supply chain through traceability data and insights.

**4. Empowering Consumer Choice:**

* We envision a future where consumers have access to detailed information about the origin, ingredients, and production processes of the food they buy. This empowers consumers to make informed choices that align with their values and dietary preferences.

**5. Strengthening Trust and Accountability:**

* Our project seeks to build trust and accountability across the entire food supply chain. We will achieve this by implementing traceability standards and technologies that ensure transparency and traceability from farm to fork.

**6. Industry Collaboration:**

* We recognize that achieving our vision requires collaboration across the food industry, including farmers, processors, distributors, retailers, and technology providers. Our project will facilitate industry-wide cooperation to establish uniform traceability standards and practices.

**7. Leveraging Technology:**

* We aim to harness the power of cutting-edge technology, including blockchain, IoT (Internet of Things), and data analytics, to create a seamless and efficient food traceability system that can adapt to evolving industry needs.

**8. Regulatory Compliance:**

* Our project vision includes ensuring compliance with food safety regulations and international standards. We will work closely with regulatory bodies to align our traceability system with existing and emerging requirements.

**9. Global Reach:**

* We aspire to create a food traceability system that is globally accessible and can be adopted by food supply chains worldwide, promoting international cooperation and consistency in traceability practices.

**10. Continuous Improvement:**

* We commit to continuous improvement in food traceability practices, technologies, and standards. We will adapt to emerging challenges and opportunities to ensure the ongoing success of our project.

**11. Enhance Transparency:**

* We will create a digital ecosystem that allows consumers to trace the journey of their food from farm to table. Through a user-friendly interface, they will access detailed information about the origin, processing, and distribution of every product they purchase.

**12. Promote Sustainability :**

* We will support sustainable food production and distribution practices by providing data and insights on the environmental impact of food products. This will enable consumers to make informed choices that align with their values, and producers to optimize their processes for sustainability.

**13. streamline Supply Chains:**

* Our technology will streamline supply chain operations, reducing inefficiencies and minimizing food waste. Producers, distributors, and retailers will benefit from improved inventory management, reduced spoilage, and increased profitability.

**14. Enable Compliance:**

* We will simplify regulatory compliance by offering a centralized platform for recording and sharing critical information. This will help food businesses adhere to international and local regulations while reducing the administrative burden.

**15. Foster Innovation:**

* We will continuously innovate and leverage emerging technologies such as blockchain, IoT, and AI to enhance traceability and data accuracy. Our open API ecosystem will encourage third-party developers to create valuable integrations and applications.

**16. Empower Stakeholders:**

* Our product will empower individuals and organizations across the food supply chain, from small-scale farmers to multinational corporations, to make data-driven decisions that improve food quality, safety, and sustainability.

Our project vision for food traceability is centered on safety, quality, sustainability, transparency, and collaboration. By achieving these objectives, we aim to create a food supply chain that benefits consumers, supports sustainable practices, and builds trust across the industry.By realizing this vision, we aim to create a world where everyone can confidently enjoy safe, nutritious, and sustainably sourced food, while promoting ethical practices and minimizing the environmental impact of the food industry. Together, we can build a brighter future for “food traceability."

**Section 2 : Datasets**

## **2.1** **Overview**

Data about Critical Tracking Events (CTEs) and the Key Data Elements (KDEs) that go with them are contained in the Traceability Rule. This dataset includes a variety of CTEs, including initial packaging, cooling, and harvesting. At particular points in the supply chain, including the initial packing step, the food product is given a special code known as the Traceability Lot Code (TLC). At any stage in the supply chain, this TLC acts as a unique identification for a specific commodity.The column names in the dataset will be renamed for better understanding and reference in order to increase simplicity and clarity.

A limited amount of rows from the sample dataset are used to illustrate various supply chain scenarios. Although there are KDEs for 10 different CTEs in this dataset, the precise CTEs can change depending on the kind of meal being tracked. Finding the pertinent CTEs for the foods in the Food Traceability List (FTL) is the goal. Not every meal listed in the FTL experiences every CTE. The Food Traceability List includes items like shell eggs, fresh tomatoes, nut butters, and leafy greens as examples.

## **2.2** **Field description**

The Food Traceability Rule requires those who produce, process, pack, or hold foods listed on the Food Traceability List (FTL) to keep track of and give their supply chain partners specific data (known as key data elements, or KDEs) for a number of critical tracking events (CTEs) in the food's supply chain. The picture below displays every significant tracking event in the supply chain, including harvesting, cooling, first packing, first land-based receiver, shipping, receipt, and transformation. This framework provides the FDA with all the information required to set up food traceability and serves as the cornerstone for successful and efficient food tracing across the supply chain.

*Figure 2****.****. CTEs as defined by the Food Traceability Rule*

| **Critical Tracking Event (CTE)** | **Description** |
| --- | --- |
| Harvesting | The product (such as fruits, vegetables, or grains) are first procured from its source (such as a farm, orchard, or field). It is customary to keep track of details like the type of product, the quantity gathered, and the date, time, and location of the harvesting. |
| Cooling (before initial packing) | To preserve freshness and avoid spoiling, products frequently need to be chilled after harvesting. This could entail techniques like hydrocooling, forced air cooling, and room cooling. It is typical to keep track of the cooling process, time, and temperature. |
| Initial Packaging | The first packaging for shipment of the harvested goods occurs at this stage. It is possible to keep track of information like the type of packing supplies used, the quantity per package, the day and time of packing, and the beginning details of transportation. |
| First Land-Based Receiver | After being transported from the harvesting site, this is the first location where the packaged goods are received. It is possible to keep track of information like the date and time of delivery, the product's state at delivery, and the storage conditions. |
| Shipping | Transporting the product from one place to another is involved here. Typically, information is kept on the shipping date and time, mode of transportation, destination, and anticipated arrival time. |
| Receiving | After shipment, the product is now received at this stage. It is possible to keep track of information like the date and time of delivery, the product's state at delivery, and the storage conditions. |
| Transformation | This includes any operation that materially modifies the product, such as cooking, dicing, slicing, combining, etc. Typically, information on the transformation process, the time and date, and the result is recorded. |
| Traceability Plan | This strategy, which describes how each CTE will be tracked and recorded, rather than a CTE, is what this is. It comprises information on the techniques and tools used for traceability, roles for gathering and storing data, and protocols for viewing and exchanging traceability data. |

## **2.3** **Data Context**

In response to how challenging and time-consuming it is to identify the source of food contamination in convoluted supply chains, the FDA has developed a food traceability rule. Beyond what is already required by law, this rule requires businesses and organizations working in the food industry to keep documents pertaining to food traceability. A variety of enterprises, including businesses, farms, retail food establishments, and restaurants, are subject to this obligation.

According to this rule, these organizations are required to provide information to the FDA within 24 hours or within a mutually reasonable time that the FDA determines. The manufacturer or seller who is in charge of using or acquiring transported food goods is required to keep track of the labeling.

**APPENDIX**

**Appendix A: Glossary**

| **Term** | **Definition** |
| --- | --- |
|  |  |
| **Foodborne illnesses** | Diseases caused by consuming food or drink contaminated with harmful germs or chemicals. |
| **Foodsupply chains** | The network of all the individuals, organizations, resources, activities, and technology involved in the production and sale of a food product, from the delivery of raw materials from the supplier to the manufacturer, through to its eventual delivery to the end consumer. |
| **Transparency** | In the context of food supply chains, transparency refers to the visibility and accessibility of information at all stages of the food supply chain. |
| **Traceability** | The ability to track and trace a food product's path within the supply chain. |
| **Food Traceability Rule** | A rule established by the FDA as part of the Food Safety Modernization Act (FSMA) that mandates additional record-keeping requirements for those who manufacture, process, pack, or hold foods included on the Food Traceability List (FTL). |

**Appendix B: GitHub Repository**

**Overview**

Food traceability in the supply chain refers to the ability to track and trace the movement of food products from their source to the final consumer. This process is crucial for ensuring food safety, quality, and transparency.

Here's an overview of critical aspects of food traceability in the supply chain: Traceability Systems: Food traceability systems use various technologies such as barcodes, QR codes, RFID (Radio-Frequency Identification), and blockchain to create a digital record of a food product's journey. Each item or batch is assigned a unique identifier that can be scanned or recorded at various points in the supply chain. Traceability Steps: The food traceability process typically involves several steps: Forward tracking

* Data Collection: Product information such as origin, production date, and batch number are recorded.
* Recording Movement: The data is updated whenever the product changes hands or places.
* Data Storage: This information is safely saved and can be accessed for tracking and verification.
* Accessibility :Consumers, regulators, and stakeholders can gain access to this information via labels, applications, or websites.

Benefits:

* Food safety: Rapid detection of tainted or dangerous items allows for prompt recalls, lowering health risks.
* Quality assurance is achieved by monitoring the conditions and handling of products.
* Transparency: Consumers can make educated decisions based on the origin and characteristics of the product.
* Supply Chain Efficiency: Traceability increases supply chain visibility while decreasing waste.
* Regulations: Many countries have traceability of food regulation and standards in place. These rules frequently outline what information must be recorded and how it must be shared.
* Blockchain: Due to its openness and security, blockchain technology is gaining favor for food traceability. It generates an immutable ledger of transactions, making data tampering or falsification difficult.
* Consumer Engagement: As customers become more interested in the origins of their food, they are turning to traceability systems to obtain insight into the items they buy. This can help customers and food producers create trust.

Implementing food traceability systems can be complicated and costly. It necessitates agreement from all supply chain players. Data security and privacy are also major issues.

Traceability becomes more important as the global food supply chain becomes more complex. Traceability systems aid in the management of international trade and the compliance with various standards.

In a nutshell food traceability in the supply chain is an important tool for assuring food safety, quality, and transparency. It entails the use of technology, data collecting, and stakeholder collaboration to follow the route of food items from farm to fork. By increasing trust and safety, it benefits both customers and the food business.

**GitHub Repository Link**

[**https://github.com/vaishnavirao05/Healthy-FoodNet**](https://github.com/vaishnavirao05/Healthy-FoodNet)

**GitHub Repository Contents**

The GitHub repository functioned as a working directory and centralized platform for all project participants to collaborate. It contains preemptive data to assist the team to get insights on the requirements. The README file offers a brief summary of the project's contents as well as step-by-step instructions to assist the client and other interested parties in using the project.

**· README.md**

**Appendix C: Risks**

**Sprint 1 Risks**

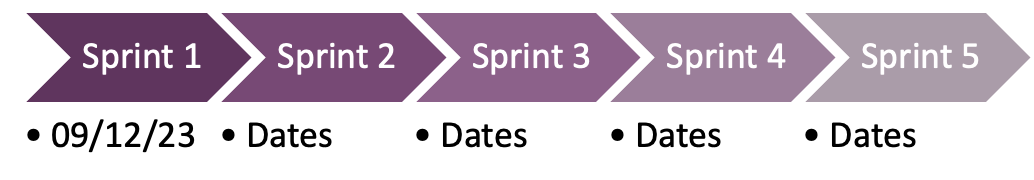
| Risk Name | Description | Probability | Impact | Mitigation |
| --- | --- | --- | --- | --- |
| Supply Chain Data | Due to a lack of data in the supply chain, it may be necessary to develop data that is artificial. | medium | medium | Incorporate real-world data and limit the use of inaccurate information. |

**Sprint 2 Risks**

| Risk Name | Description | Probability | Impact | Mitigation |
| --- | --- | --- | --- | --- |
| **Mapping CTE Events** | **Mapping each CTE event to next event to form supply chain** | **High** | **High** | **Focus on small data sample, construct end-to-end supply chain** |

**Appendix D: Agile Development**

Scrum Methodology



Sprint 1 Analysis

During this sprint, our primary objectives revolved around formulating a clear problem statement and generating an initial report outlining the project's goals and challenges. We've gained valuable insights throughout this process, which can provide valuable guidance for future teams embarking on similar projects.

One notable aspect of our sprint was the effective collaboration among team members, each bringing their unique expertise in data science and software development to the table. This diversity allowed us to approach the task from multiple angles, fostering creativity and innovation. We prioritized transparent communication, establishing clear channels for discussions and holding regular team meetings to review progress, address questions, and resolve any ambiguities that arose. The use of project management tools was instrumental in helping us monitor tasks, assign responsibilities, and ensure the project remained on track according to our established timeline.

Reflecting on our experience, we recognize that we could have improved our process by conducting more comprehensive reviews of each other's work rather than merely relying on quick proofreads. To streamline our efforts and avoid duplicating work, we established a central GitHub Repository where we compiled our findings and documents.

Overall, Sprint 1 served as a pivotal phase in laying the groundwork for our project. The lessons learned, particularly regarding effective collaboration, clear communication, and efficient project management, have set us on a strong path moving forward.

**REFERENCES:**

[1]. Center for Food Safety and Applied Nutrition. (n.d.-a). *Investigations of foodborne illness outbreaks*. U.S. Food and Drug Administration. <https://www.fda.gov/food/outbreaks-foodborne-illness/investigations-foodborne-illness-outbreaks>

[2]. Center for Food Safety and Applied Nutrition. (n.d.-b). *Outbreaks of foodborne illness*. U.S. Food and Drug Administration. <https://www.fda.gov/food/recalls-outbreaks-emergencies/outbreaks-foodborne-illness>

[3]. Centers for Disease Control and Prevention. (2023, August 31). *Foodborne outbreaks*. Centers for Disease Control and Prevention. <https://www.cdc.gov/foodsafety/outbreaks/index.html>

[4]. Communicable disease report for healthcare providers. (n.d.-a). <https://www.azdhs.gov/documents/preparedness/epidemiology-disease-control/disease-investigation-resources/communicable-disease-report-form.pdf>

[5]. *Foodborne disease outbreaks*. Environmental Health. (2017, February 24).<https://www.vdh.virginia.gov/environmental-health/food-safety-in-virginia/foodborne-diseases-and-outbreaks/foodborne-outbreaks/>

[6]. introduction to foodborne illness outbreak investigations - books. (n.d.-b). <https://ajph.aphapublications.org/doi/full/10.2105/9780875532943ch13>

[7]. Research on the application of big data in smart food Safety. (2020, September 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/9546382>

[8]. Study on Microwave Food Contamination Examination based on Machine Learning approach. (2022, December 11). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/10015961>