**Assessing Future Availability Risk in Food Supply Chains**

**SupplySight - Team 16**

**Create X Capstone Design CS 4723-X01**


-

**Instructor: Craig R Forest**

**Mentor: Daniel H Forsyth**

**Georgia Institute of Technology**

**Atlanta, GA 30332**

-

**Valentina Garcia**

**Vasili Fovos (Primary Editor)**

**Jinho Lee**

**Chenyue Wang**

**Hongyi Wang**

**Table of Contents**

**Executive Summary**

Small to mid size food distributors in the U.S. face persistent risk regarding future product availability particularly for internationally sourced goods but not excluding national products. This risk often forces purchasing decisions to be made without the necessary data, relying mainly on spreadsheets, supplier conversations, and intuition from years of experience. As global supply chains become increasingly more volatile due to climate change, geopolitical uncertainty, trade restrictions, and the general post-pandemic nature of economies, the financial burden caused by incorrectly stocking products has become profound. Despite this widespread challenge, enterprise supply chain systems primarily focus on demand and network level risk rather than shorter term product level prediction.

To evaluate the validity of this problem, our team conducted many customary discovery interviews with small to mid size food distributors. From these interviews, 75% of the distributors identified future product availability as their primary purchasing challenge which occurs weekly to bi weekly in line with order cycles. This places impactful financial stress on these companies in the range of 1000 to 10000 dollars based on the severity of the shortage. This confirms the validity of the problem and the impact it causes, making a clear opportunity within this specific market segment for a viable solution.

The proposed solution for this problem is a predictive product availability risk platform for small to mid size food distributors. This system will aggregate public data including agricultural reports, weather patterns, product import data, and finally industry news to create risk scores for monthly periods. In order to make predictive output based on this data, we will apply a supervised Random Forest model which was seen as the best for supply chain risk prediction (Xu et al., 2024). These risk predictions will be presented through an interactive dashboard that allows for product filtering. Furthermore, this platform will explain the reasoning for the conclusion based on the data and news that corresponds to the risk analysis. To understand the success of our platform, we will use metrics that include probabilistic classification accuracy greater than 75%, inference time less than 5 seconds, monthly model updates, and clear risk categorization.

In order to achieve this solution, we have many concept ideations including rule-based scoring systems, supervised machine learning architectures, and hybrid dashboard-chatbot interface. Among these options, we choose a supervised Random Forest based dashboard system as the preferred architecture due to the many constraints of success and functionality requirements. Furthermore, the academic literature on supply chain risk aligns with this model choice due to its predictive strength with non-linear interactions.

To date, the team has conducted customer interviews, defined functional requirements, completed landscape analysis, and generated preliminary system architecture concepts.

The next steps for our platform will focus on engineering data pipelines, training and validating our predictive model, backtesting using defined historical supply chain disruptions, and deploying a functional platform interface. We would also like to have user testing to determine possible changes to our platform prior to final implementation. The long term goal of this product is to provide a data driven tool used by food distributors that improves decision and financial stability when related to future product availability. This will allow small to mid size food distributors operate more efficiently through uncertainty in the volatile global supply chain.

## Nomenclature

API – Application Programming Interface

LLM – Large Language Model

AUC-ROC – Area Under the Receiver Operating Characteristic Curve

ETL – Extract, Transform, Load

Risk Score – Predicted probability of product shortage

Cron Job – Scheduled automated task

## 1.0 Introduction and Background

### 1.1 Problem, Motivation, and Need

The Food Supply Chain is an extremely unpredictable environment. They frequently have little insight into potential disruptions in supply and demand, so it's hard to make decisions about what to buy, how much to buy, and when to buy it. These decisions carry a large financial risk because food products are perishable. Margin loss, waste, spoiling, stockouts, lost sales can result from underestimating and overestimating demand.

The lack of proactive tools to detect potential availability risks in the food supply chain is the design issue this project attempts to address. Future availability risk is the likelihood that a product will become insufficient or unavailable soon as a result of external events (e.g., weather or geopolitical events), seasonality, price volatility, supplier delays, demand volatility.

Many distributors today rely on spreadsheets, ERP systems, historical sales data, and human intuition. Although these tools facilitate simple forecasting, they don't offer early warnings of new threats. With a predictive system that anticipates possible availability disruptions, distributor managers will conduct informed planning and make purchasing decisions more efficiently.

### 1.2 Customer Discovery

We did early-stage customer discovery by finding and contacting food distributor managers, buyers, and supply chain analysts at food distributors to learn more about this problem. The purpose of these interviews was to clarify:

- How people decide what to buy right now
- What information is used to make predictions
- When there are gaps in information
- What operational or financial problems can result from missing information

Initial research indicates that:

- The margins on food distribution inventories are small, which means small mistakes make a big impact.

- There is a lot of volatility in categories like seafood and fresh fruits and vegetables.
- Forecasting systems frequently do not incorporate external risk signals.

### 1.3 Technical issues, Challenges and Opportunities

During the development of this system, we can encounter problems such as:

- Data integration: Combining internal sales data with external signals like weather, price indices, and supplier reliability metrics.
- Data quality: Historical data that is not consistent or complete.
- Forecast uncertainty: Demand and supply shocks are inherently random.
- Explainability: Procurement managers need to know why a risk flag goes up.
- Cold-start problem: There isn't any historical data on new products.

### 1.4 Desired solution

Despite the technical issues that can emerge, our desired solution focuses on predicting availability risks across the food supply chain using probabilistic models and structuring risk scoring. The system combines machine learning with external data sources to improve prediction accuracy over time. These insights are then translated into a dashboard highlighting potential risks and allowing stakeholders to make more informed decisions.

## 2.0 Teams and founders

### 2.1 Tasks/Roles:

Valentina Alves De Aquino Garcia:

- collect news feeds/APIs
- design LLM prompts
- define JSON schema
- validate outputs
- build ingestion pipeline

Vasili Gregory Fovos:

- design DB schema
- connect ETL + LLM + ML outputs
- build API to serve risk scores
- ensure data consistency
- performance optimization

Jinho Lee:

- find datasets/APIs
- clean and normalize data
- create ingestion scripts
- store in DB schema
- schedule updates

Chenyue Wang:

- define target variable
- build feature set
- train models
- evaluate performance

- design scoring logic

Hongyi Wang

- design dashboard structure
- decide what stakeholders see
- build frontend
- connect backend APIs

**2.2 Commitment level**

Our goal is to create a working product that fits with our Minimal Viable Product vision by the end of the semester. We don't plan to keep working on the project after the course is over, but we do want to focus on making something we are really proud of that shows we can design and build a real solution. For us, success means getting real-world experience with machine learning, learning how technical ideas become products, and experiencing a startup-style setting. We also think it's important to be able to show our work to professors and guest experts and learn from their feedback, ideas, and experience in the field.

**2.3 Major Decision**

For major decisions we aim to reach a shared agreement so that everyone's opinions are taken into consideration. We value an open communication and open mindset which is crucial to create an environment where members feel comfortable contributing ideas and feedback. We also follow agile principles by regularly sharing updates so the team stays aligned and accountable throughout the project.

**2.4 Solution if Co-founder "Loses Interest"**

In case of a co-founder loses interest:

- Responsibilities will be redistributed based on skill alignment.
- Remaining team members will reassess scope and reduce feature set if necessary.
- Open communication will be prioritized to avoid abrupt disengagement.

Since this is an academic project, continuation beyond semester would be voluntary and mutually agreed upon. If we form a startup beyond the semester, and after we share the equity equally a co-founder disengages:

- They retain only vested equity.

- Unvested shares return to the company.
- Role responsibilities are reassigned.
- The remaining founders may reallocate returned shares.

If departure is harmful or sudden, buyback provisions may apply at fair market value.

**2.5 Team After Semester**

After the semester, anyone interested will continue the development.Members may contribute part-time if aligned with academic or career goals. If no continuation occurs, the project will remain a completed academic prototype. For now the team's top priority is to produce a credible and well-engineered system during the course timeline. If a food distributor uses our product during the semester, we would clearly communicate that the platform is a prototype for evaluation and feedback only and should not be solely relied upon for operational decisions. We would monitor the system at a basic level and fix critical issues when feasible, but we would not guarantee uptime or long-term maintenance.

## 3.0 Business Thesis

### 3.1 Thesis

Small to mid-sized food distributors will buy a food availability intelligence platform because it gives them forward-looking confidence on future product availability and helps them decide when to buy, wait, or find alternative products.

This thesis is based on the fact that small to mid-sized food distributors' companies keep experiencing uncertainty when they order products like frozen seafood from Asia. These products are usually affected by climate change, trade environment, and supply chain disruptions. Research shows that these events increase the uncertainty in the supply chain (Rojas-Reyes et al., 2024). However, most food distributors still rely on experience-based decisions rather than evidence-based decisions.

After interviewing more than 12 food distributors. Our hypothesis was

- 75% of small to mid size food distributors had purchasing uncertainty regarding future product availability that occurs weekly or bi-weekly
- Purchasing uncertainty leads to significant financial losses due to overstocking, understocking, and emergency purchasing. Reported in the $1000 to $10000 per week.

### 3.2 Value Proposition

Food Distributors are not getting enough reliable information to make purchasing decisions. This causes inefficient inventory allocation and operational stress. They have to repeatedly decide whether to delay purchases or find substitutes without insight into future supply conditions. Customer discovery interviews confirm that 75% of food distributors identify the risk of uncertainty in future product availability as their primary purchasing issue. This leads to emergency purchasing or excess stocking which results in financial losses. Especially in internationally sourced products such as seafood or paper products. This has become extremely prominent in the last 5 years due to climate change and geopolitical disruption which has increased the risk of supply chain instability.

## 4.0 Landscape Deep Dive

### 4.1 Target Market and Customer Segmentation

The target market for this project is small to midsize food distributors in the United States. This subsection of the broader food distribution industry is worth about 40% of the total market share which is around 700 billion Dollars. This makes this market size about 300 billion dollars. Our product will target small to mid size food distributors as they are more exposed to uncertainty in future product availability due to limited access to analytical resources, lower supplier leverage, and inability to push higher costs onto customers like national distributors. Within small to mid distributors, our first step further segmentation includes limit in frozen seafood product type, owner-operated small team grocery stores who relied on spreadsheets and high sensitivity to volatility. We decided to focus on this as purchasing decisions occur more frequently than large corporations and the cost of incorrectly stocking products is significant and impactful on their revenue and cashflow.

### 4.2 Industry Trends

In the last five years, we see an increase in geopolitical instability including tariffs and trade restrictions, more frequent extreme weather events affecting agriculture, post COVID supply chain fragility awareness and increased reliance on import food products. All of these factors have created a need for data driven tools that help with forward looking insights on product availability.

### 4.3 Competitive Landscape and Differentiation

Currently, distributors rely on supplier information, historical data, and several applications. There are several companies that operate in adjacent areas of supply chain planning and risk management such as Resilinc, SAP, and RELEX solutions. These are all solutions that focus on providing ERP platforms that work in the space of demand forecasting, supply chain disruptions, and inventory management. These are effective at tracking historical data and predicting the demand side of purchasing but do not focus on the future availability of product. This creates an opportunity in the supply side of purchasing in order to help food distributors make better decisions when planning for future availability of product. It is important to note that this is not an integrated software like an ERP, IMS, or SRM, but instead a standalone app that is meant to support decision making in coordination with these services.

### 4.4 Buying Process, Pricing, Distribution, and Regulatory Considerations

Our buyers are the business owners and CFO, while users will be purchasing managers and inventory planners. Our software purchases will be annual while inventory purchasing decisions occur weekly or biweekly with yearly total evaluation. We may propose a monthly pricing at 150-500 depending on the number of categories available to predict and company size. We will distribute our software in direct sales outreach, online website distribution or partnership with other common tools that distributors use now. While our proposed tool does not directly handle regulations, we must ensure our data integrity and avoid misleading forecasts that impact macro supply chains that may cause man-made widespread shortages.

### 4.5 Current Customer Behavior and Product Positioning

Today food distributors rely mainly on conversations with suppliers, their historical purchasing data, and intuition from experience to help manage uncertainty in this space. Even when distributors have advanced ERP systems in place to do demand forecasting, these usually extract reports rather than give insights into future product availability and risk. The proposed solution is designed to be used alongside the existing software as a decision support layer and would be a monthly subscription in alignment with the current purchasing behavior of inventory management software.

## 5.0 Existing Products

### 5.1 Existing Competing Products

Expana/Mintec is the current leader of the agrifood market intelligence industry. They provide data and price trends on a wide range of commodities. They also survey experts in the industry to improve their LLM responses. Vesper is an upcoming competitor that uses AI to forecast prices. Both of their products rely on proprietary data and are only affordable by the largest food distributors. Everstream Analytics and Resilinc are supply chain risk management platforms that support network level supply chain disruption visibility. Their main customers are manufacturers but they also support monitoring for agrifood too but without any product level risk. Gro Intelligence was another agritech company that focused on using AI and public data to predict food security. Their platform supported a wide range of products but had poor PMF. They could not raise enough funds and bankrupted in 2024.

### 5.2 Competitive Benchmarking and Strategic Differentiation

After reviewing many different competitive products in the space, we were able to get both important details about the technical design and the strategic positioning for our product. None of the products above directly address short term product availability for food distributors but all have adjacent products that can inform our design strategy.

All the competitive products use multi source data ingestion, supervised learning techniques, and risk scoring based on a certain set of data of public data and news. This validates our idea to use structured food/agriculture from many different sources, use a model to predict the outcomes of this data, and finally present these results to our customers. Unlike all the existing products though, we are going to target a small subset of an industry, narrowing our scope down to product level availability analysis for small to mid market distributors. This will allow us to provide value to our customers that does not span the entire network level supply chain therefore allowing us to compete with products with more resources.

## 6.0 Customer Requirements and Engineering Design Specifications

### 6.1 Stakeholders

The stakeholders include:

- Primary Stakeholders
  - Buyers are small to mid size food distribution companies
  - Primary users are warehouse managers responsible for ordering inventory
- Secondary Stakeholders
  - Food manufacturers, food producers, or suppliers whose production influences supply signals
  - Customers of food distributors who depend on consistent product availability from food distributors like restaurants and grocery retailers
  - Team who works on project development
  - Advisors and Mentors that oversee project execution

Although the food distribution companies and warehouse managers are primary stakeholders, improving purchasing decisions affect the entire supply chain therefore affecting both companies upstream and downstream in this industry.

### 6.2 Functional Requirements

The system will perform these functions:

- Data updating and preprocessing:
  - Automatically collect and refresh external signals relevant to supply availability prediction periodically and clean the data into our format. For unstructured data like news, we need to use a model to turn long text into json data in {title, sentiment, risk, product…}. Then we map all these data into distributor interested categories.

    - Metrics: success rate > 99%, refresh latency < interval, text-data relevance > 90%, miscategory rate < 10%.
- Risk scoring:
    - Using all relevant data, we use a machine learning model to predict a risk score on future shortage of supply in the specific category in 30/60/90 days.
    - Metrics: discrimination AUC-ROC > 0.75, warning lead time > 14 days, and recall rate > 80%.
- Decision supports:
    - Convert our risk scoring into actual action guidance our user can use.
    - Metrics: acceptance rate > 70%
- Reasoning:
    - Support with several highest ranked evidence from our database with url links
    - Metrics: user rating > 80%, evidence coverage > 90%.
- Dashboard:
    - Provide a user friendly dashboard that supports customers need, including risk threshold slider, what-if simulator, manual override on data, weight adjustments, historical and future graph, model selection and etc to let users feel controlling .
    - Metrics: load time < 2 sec, user rating > 80%
- Security:
    - Ensure the data security and consistency from attackers and mistakes
    - Metrics: uptime > 99%, backup cycle < 10 days

### 6.3 Constraints

The application is using only public open source data, which is updated weekly or bi-weekly. Which means it is hard to update the result in real-time and is unable to respond to changing circumstances. Private data, such as distributor transaction data and supplier inventory feed, cannot be included in the model, so rather than predicting the inventory of individual companies, this system focuses on providing a risk indicator. In terms of accuracy, it is impossible to predict the future with 100% certainty. Since the supply chain is influenced by various factors beyond climate and historical data. The user should understand it as informational guidance. The platform does not make any purchasing decisions; users have full responsibility for any loss, and the system holds no legal liability.

### 6.4 Set of Detailed Engineering Design Specifications

Our application helps small to mid-sized food distributors make informed decisions when purchasing products at the market. We train an anomaly detection model using Scikit-learn and backtesting techniques with agricultural and weather data from sources such as USDA and NOAA to predict the likelihood of a product having a shortage and generate a risk score. We then provide possible explanations of the supply status by linking relevant events and articles from news around the world gathered using API tube. We ingest both structured data (import, temperature, anything quantifiable) and unstructured data (news) on a daily basis. When a user searches for a product using our platform, a risk score is output within 2000ms and summaries to relevant articles within 500ms. Accurately predicting the supply status requires resources and cooperations from local vendors that we currently do not have access to. Instead our MVP guides the distributors through the food purchasing process by providing relevant knowledge. We aim to achieve a risk score generation that reflects the current state-wide supply status with greater than 90% recall rate and source 5 articles that are relevant to the situation with greater than 70% accuracy based on user ratings.

### 6.5 Relative Importance of Specifications

We want our risk score to reflect the state-wide availability of a product but its accuracy is not the top concern since state-wide product availability might not be directly relevant to a specific location. Providing news relevant to the supply status is a priority because our product at this stage still requires the distributors' knowledge and accurate information can help them make better decisions. We ingest new agricultural and weather data on a daily basis to achieve a balance between reflecting up-to-date situations and managing server workload. The latency requirement is realistic and maintains decent user experience.

### 6.6 Engineering Design Specifications

- Risk score generation (anomaly detection)
    - Based on data from no earlier than the past 3 months
    - Greater than 90% accuracy based on backtesting
    - Output within 2000ms
- Relevant articles (product specific)
    - Articles published within the last 6 months
    - Greater than 70% accuracy based on user ratings

    - Output within 500ms (db query)
- Data pipeline
    - Run daily using cron job
    - Simple ETL for structured data (import, temperature, anything quantifiable)
    - LLM to json for unstructured data (article and summary)

## 7.0 Preliminary Design Concept Ideation

### 7.1 Functional Requirements

Collect open source data(USDA, weather), Normalize different data that effect supply chain, calculate risk score of each product, Explain risk factor by news, Decision support assistance

The application is to provide decision-making assistance by analyzing data from the food supply chain. The model must collect and process the open source data that influences supply chain stability. Calculate each product's risk score. In addition to providing the risk scores, the system explains the key contributing factors, including relevant news and events. Finally, the platform offers decision support assistance by presenting structured guidance to help users consider a purchasing strategy.

### 7.2 Preliminary Concept Generation

#### 7.2.1 Function 1: Data Ingestion

- Manual Data set upload
- Automated API based ingestion from public data sources
- Hybrid approach that allows for both types for flexibility
- Most likely selected: Automated API ingestion due to its scalability and reduction of manual error.

Most likely selected: Automated API ingestion due to its scalability and reduction of manual error.

#### 7.2.2 Function 2: Risk Prediction Method

- Logistic regression learning model
- Random Forest learning model
- Rule-Based weighted scoring system

Most likely selected: Random forest learning model due to its strong predictive accuracy with non linear relationships. Many academic papers on supply chain models indicate this was the best model for predictive accuracy when integrating multi source disruption signals supports our functionality requirements (Xu et al., 2024).

#### 7.2.3 Function 3: Risk Interpretation

- Risk categorization (Low, Medium, High)
- Probabilistic output
- Risk category with top driving factors

Most likely selected: Risk category with top driving factors to improve user ability and trust in the model. Outputs Risk, articles, and data that align with the prediction.

#### 7.2.3 Function 4: User interaction

- Static dashboard
- Interactive dashboard with filtering
- Hybrid Dashboard and chatbot system

Most likely selected: Hybrid interaction dashboard and chatbot system allow users to better understand how data is supporting the risk output along with user friendly dashboard.

### 7.3 Integrated Concept Development and Evaluation

Using the selected functional solutions, three integrated system-level concepts were derived by combining data ingestion, modeling, interpretation, and user interaction layers into complete architectures.

#### 7.3.1 Integrated Concept 1 – Machine Learning Dashboard System

Features of this concept:

- Automated API-based data ingestion
- Random Forest supervised learning model
- Risk category + top driving factors output

- Interactive dashboard interface

In this concept, our platform will ingest public trade, climate, and other agriculture volatility data in order to generate availability risk probabilities. These probabilities will then be put into categorical risk levels, and presented through a dashboard where users can filter by product.

#### 7.3.2 Integrated Concept 2 – Machine Learning + Hybrid Dashboard/Chatbot System

Features of this concept:

- Automated API-based data ingestion
- Random Forest supervised learning model
- Risk category + top driving factors output
- Interactive dashboard interface
- Conversational chatbot interface

In this concept, our platform will do all the same functional things as concept 1, but we will add a chat system where users can ask questions to gain more insight about the risk outputs. This will give users more confidence and trust in the data and reasoning our model is using.

#### 7.3.3 Integrated Concept 3 - Lightweight deterministic Risk Platform

Features of this concept:

- Automated API-based ingestion
- Rule-based weighted scoring
- Risk category output
- Static dashboard

This concept eliminates machine learning and relies on predefined feature weights to compute composite risk scores. This would be the most simple and highly feasible but, it lacks a learning model that can adapt to nonlinear interactions among supply indicators.

### 7.4 Preliminary Feasibility Analysis

#### 7.4.1 Technical Feasibility

Concepts 1 and 2 are technically feasible given that:

- Available structured public data (USDA, NOAA, import data, and news sources)
- Established machine learning libraries for our use cases supported by research (Xu et al., 2024) Defined prediction horizons with known shortage times for training data
- 8-week development timeline with 5 team members who are all computer science majors

#### 7.4.2 Operational Feasibility

The proposed system does not require any integration with the existing systems that food distributors already use for inventory management. This means that our platform can stand independent from their other software and act as a separate decision support layer which companies use in collaboration with existing systems.

#### 7.4.3 Resource feasibility

Development workload distribution:

- Data pipeline engineering
- Feature engineering and labeling
- Model development and validation
- Dashboard front-end development
- Integration and testing

The scope is manageable within the semester for five developers

#### 7.4.4 Preliminary Risks and Countermeasures

- User trust risk:

If the evidence is not explained, users may not feel reliable about the risk score. Simple numerical output can limit operational use. In the B2B decision-making process, the user's trust affects system reliance. It needs to provide interpretable insights to support managerial decision-making (Baryannis et al., 2019)

Countermeasures: The system describes the main factors that influence risk score. It also displays related news articles and data along with visualizations. This enhances transparency.

- Data update risk:

  Public data sources such as USDA reports, NOAA climate data, and trade statistics are updated at different frequencies. Different update cycles can reduce responsiveness.

  Countermeasures: System will use time alignment logic to calibrate the update cycle. Also it will display the timestamp of the last update for transparency.

- Project schedule risk:

  8 weeks might not be enough time to implement all the features. (machine learning pipelines, data engineering processes, dashboard interfaces, and chatbot functionality) It increase the risk or incomplete integration.

  Countermeasures: Using phased prioritization strategy. A minimal functional pipeline(data engineering processes, dashboard interfaces) will be implemented first for a working prototype. Advanced features(chatbot) will be a stretch goal. It will decrease the schedule risk.

- External dependency risk:

  As most of the data is from external sources, API changes, API failure, API error, or access restriction can occur. It makes the system unstable, which leads to degraded user experience and lower reliability.

  Countermeasures: The system will store all processed data into an internal database to establish a fallback dataset. If an API failure occurs, the system will calculate a risk score based on collected data. Additionally, a local caching mechanism will be implemented to deal with temporary API errors. Modular Architecture will be used so that even if the API changes, this will allow easy modification.

#### 7.4.5 Preliminary discussion

Design selection will be discussed based on three criteria: Predictive reliability, explainability, implementation feasibility.

- Predictive reliability:

  This model aims to make the user's decision process easier. If the risk score is inaccurate, the meaning of decision support will disappear, and it will only serve as a visualization or data collection, which is completely different from the system we intend.

- Explainability:

  Our main target group, food distributors, do not trust unless they understand what data it is based on. In particular, they have been making intuitive decisions for a long time, so sufficient explanation is needed to attract the main consumer base.

- Implementation feasibility:

  Since the prototype is needed in 8 weeks, it is also very important to determine whether 5 CS major students can implement it.

At this stage, Concept 1 is considered as the most appropriate design direction. In terms of predictive reliability, the Random Forest model can effectively capture flexible supply chain risks such as climate change (Breiman, 2001). In terms of explainability, it presents a top driving factor. Also, it is considered doable in 8 weeks. Concept 2 is the best performance, but additional features such as a chatbot make it less feasible, and concept 3 is the easiest, but it is less accurate because it only performs simple scoring.

**Summary**

Through several interviews with managers of food distributors, restaurant owners, and brokers, it was identified that various factors, such as changes in the trade environment and climate change, have a significant impact on the supply stability of foreign products, including Asian frozen seafood, resulting in periodic uncertainty about product supply.  Although this uncertainty has a significant impact on purchasing decisions, most of the current decisions do not use practical data or evidence-based prediction tools. Most decisions rely on personal experience or relationship-based information. Analysis of existing platforms Expana and Vesper indicates that they are focusing on high-priced solutions, revealing a market gap in decision-making support tools for small and medium-sized distributors.

Based on this, major functional requirements were defined: data collection, risk prediction modeling, decision support, real-time update, and explanation of major factors. Now, the structure that combines a Random Forest-based prediction model with an interactive dashboard is considered the most likely design direction.

Plans focus on technology implementation. As a short-term plan, a data collection pipeline and database will be developed using reliable public data to prepare a dataset for model learning. In addition, tests will be conducted to determine the most suitable prediction engine.  In the long term, a prototype capable of actual consumer testing will be implemented, and the system will be improved based on collecting feedback from actual industrial stakeholders.

Maintain continuous communication with stakeholders in order for the direction of development to align with the needs of the field. Progress and goals are reported to the mentor every week to monitor the integration status and technical challenges. If there is issue or guidance is needed the team will contact directly. Internally team shares information every day, updates the development schedule, and implements tasks of the project. The basic schedule follows the Figure 1 and Figure 2. At major milestones, such as final design confirmation or prototype completion, the team will monitor project with previously interviewed industry stakeholders and collect feedback. Communication will extend beyond food distributor managers to include brokers and restaurant owners as indirect stakeholders. Through iterative feedback cycles and continuous testing, collected information will be documented and used in the development process.

## Gantt Chart

| Task | Start Date | End Date | Duration |
|---|---|---|---|
| Define risk scope and success criteria | Feb 17 | Feb 19 | 3 |
| Choose datasets (USDA, weather, news) | Feb 18 | Feb 21 | 4 |
| Design pipeline architecture diagram | Feb 20 | Feb 24 | 5 |
| Design database schema | Feb 21 | Feb 25 | 5 |
| Define LLM JSON output format | February 24 | Feb 26 | 3 |
| Build USDA ingestion pipeline | March 1 | Mar 8 | 8 |
| Build weather ingestion pipeline | March 3 | Mar 10 | 8 |
| Set up news data collection | March 5 | Mar 11 | 7 |
| Implement LLM extraction for news events | March 8 | Mar 15 | 8 |
| Store structured + extracted data in DB | March 10 | Mar 16 | 7 |
| Define features for risk model | March 18 | Mar 21 | 4 |
| Prepare training dataset from DB | March 20 | Mar 24 | 5 |
| Train first ML model prototype | March 25 | Mar 30 | 6 |
| Evaluate model and refine features | March 25 | Apr 3 | 6 |
| Implement risk score output logic | April 2 | Apr 5 | 4 |
| Connect model to database pipeline | April 4 | Apr 7 | 4 |
| Build API or script to serve predictions | April 6 | Apr 9 | 4 |
| Test end-to-end pipeline | April 8 | Apr 11 | 4 |
| Design dashboard layout | Apr 10 | Apr 12 | 3 |
| Build dashboard frontend | Apr 12 | Apr 18 | 7 |
| Connect dashboard to backend/API | Apr 15 | Apr 20 | 5 |
| Add alerts and explanation features | Apr 18 | Apr 22 | 4 |
| System testing and debugging | Apr 18 | Apr 23 | 6 |
| Prepare presentation slides | Apr 20 | Apr 23 | 4 |
| Record demo backup/screenshots | Apr 22 | Apr 24 | 2 |
| Final rehearsal and polish | Apr 23 | Apr 25 | 2 |

Figure 1: Gantt Chart in table

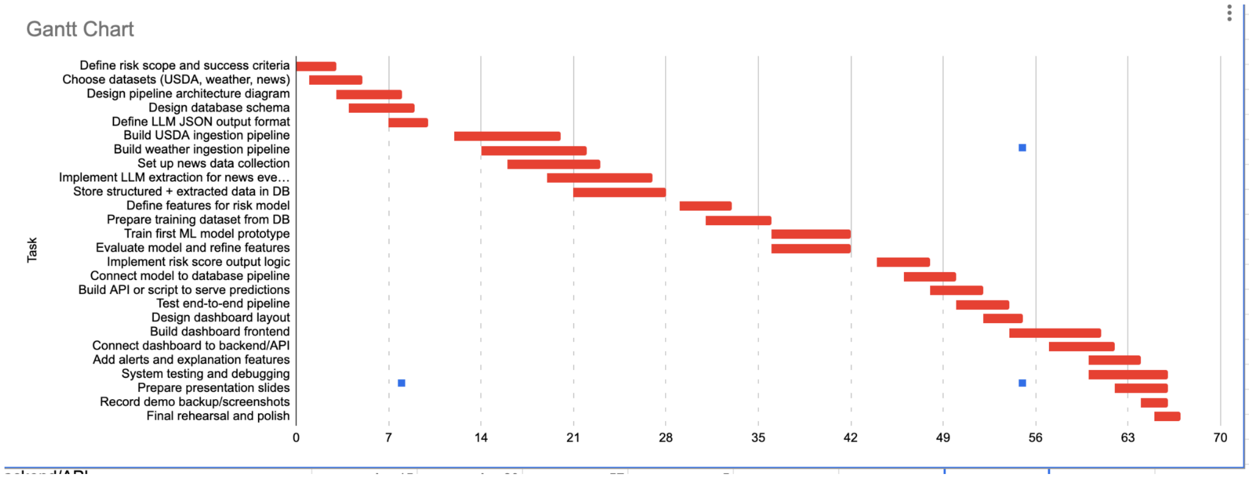


Figure 2: Gantt Chart

**References**


Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. International Journal of Production Research, 57(7), 2179–2202. https://doi.org/10.1080/00207543.2018.1530476

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

McKinsey & Company. (2022). *Risk, resilience, and rebalancing in global value chains*. https://www.mckinsey.com/capabilities/operations/our-insights/risk-resilience-and-rebalancing-in-global-value-chains

Resilinc. (2023). *Supply chain risk management solutions*. https://www.resilinc.com

Rojas-Reyes, J. J., Rivera-Cadavid, L., & Peña-Orozco, D. L. (2024). Disruptions in the food supply chain: A literature review. *Heliyon*, *10*(14), e34730. https://doi.org/10.1016/j.heliyon.2024.e34730

United States Department of Agriculture. (2023). *Economic Research Service: Data and reports*. U.S. Department of Agriculture. https://www.ers.usda.gov

United States Department of Agriculture. (2023b). *Food service industry: Market segments*. U.S. Department of Agriculture. https://www.ers.usda.gov/topics/food-markets-prices/food-service-industry/market-segments

United States Department of Agriculture, Agricultural Marketing Service. (2023). *Agrifood supply chain report*. U.S. Department of Agriculture. https://www.ams.usda.gov/sites/default/files/media/USDAAgriFoodSupplyChainReport.pdf

United States International Trade Commission. (2023). *DataWeb: U.S. import and export statistics*. https://dataweb.usitc.gov/

Xu, Z., Ni, W., Liu, S., Wang, F., & Li, J. (2024). Logistics supply chain network risk prediction model based on intelligent random forest model. *IEEE Transactions on Engineering Management, 71*, 9813–9825. https://doi.org/10.1109/TEM.2023.3317922

National Oceanic and Atmospheric Administration. (2023). *Climate data and weather anomalies*. U.S. Department of Commerce. https://www.noaa.gov

**Appendices**

Appendix A – Academic Support for Random Forest Model Selection

This appendix summarizes the academic foundation supporting the selection of a Random Forest based supervised learning model for supply chain risk prediction.

Xu et al. (2024) develop an Intelligent Random Forest model for logistics supply chain network risk prediction under uncertainty. The study integrates multi-source structured data and demonstrates strong predictive performance in classification tasks involving heterogeneous risk factors. The authors show that ensemble tree-based methods are well suited for modeling nonlinear relationships and complex feature interactions within supply chain systems.