INNOVATIONS IN TECHNOLOGY AND SCIENCE EDUCATION
ISSN 2181-371X
https://sr-journals.org/index.php/ITSE

AI POWERED INVENTORY MANAGEMENT SYSTEM FOR GROCERY STORES

Mirjalol Karimov, *BSc (Hons) Business Information Systems, Level 6, Westminster International University in Tashkent*
Pooja, *Senior Lecturer, Computing Department, School of Law, Technology and Education, Westminster International University in Tashkent*

Abstract Effective inventory management enables grocery stores to prevent stockouts, reduce waste, and allocate resources more efficiently. This project looks at the forecasting problems Korzinka faces. Korzinka is Uzbekistan's biggest modern grocery retailer, with over 190 stores in three formats and 94 million customer transactions each year. The company's 2024 Annual Report shows a product availability rate of 92%, which is below the industry standard of 95%. Shrinkage rates in proximity stores are 4.5%. Korzinka also plans to grow to 1,000 stores by 2028. At that size, manual forecasting will not be enough. The project came up with a possible solution with an AI-driven demand forecasting system by implementing and comparing three machine learning algorithms: Linear Regression (baseline), ARIMA (time series), and Random Forest (ensemble). Using the dataset from Kaggle - Store Item Demand (913,000 records, 10 stores, 50 items, 5-year period), a supervised learning pipeline with 20 engineered features was constructed. Random Forest delivered the best results (MAPE: 13.35%, R^2: 0.922), outperforming Linear Regression (MAPE: 15.66%) and ARIMA (MAPE: 27.10%). A web dashboard and stock recommendation engine was developed and deployed to a server to convert forecasts into actions. The results show that ensemble methods significantly improve forecasting accuracy for multi-store retail, with an estimated annual impact of UZS 150–200 billion for Korzinka.

Abstract. Samarali inventarizatsiya boshqaruvi oziq-ovqat do'konlariga zaxiralarning tugashini oldini olish, chiqindilarni kamaytirish va resurslarni samaraliroq taqsimlash imkonini beradi. Ushbu loyiha Korzinka duch keladigan prognozlash muammolarini ko'rib chiqadi. Korzinka O'zbekistonning eng yirik zamonaviy oziq-ovqat chakana sotuvchisi bo'lib, har yili uchta formatda 190 dan ortiq do'konga va 94 million mijoz tranzaksiyalariga ega. Kompaniyaning 2024-yilgi yillik hisobotida mahsulot mavjudligi darajasi 92% ni tashkil etadi, bu sanoat standarti 95% dan past. Yaqin atrofdagi do'konlarda qisqarish darajasi 4,5% ni tashkil qiladi. Korzinka shuningdek, 2028-yilga kelib 1000 ta do'konga yetishni rejalashtirmoqda. Bu hajmda qo'lda prognozlash yetarli bo'lmaydi. Loyiha uchta mashina o'rganish algoritmlarini: Chiziqli regressiya (bazaviy), ARIMA (vaqt qatorlari) va Tasodifiy o'rmon (ansambl) ni joriy etish va taqqoslash orqali sun'iy intellektga asoslangan talabni prognozlash tizimi bilan mumkin bo'lgan yechimni topdi. Kaggle - Store Item Demand (913 000 ta yozuv, 10 ta do'kon, 50 ta buyum, 5 yillik davr) ma'lumotlar to'plamidan foydalangan holda, 20 ta muhandislik xususiyatlariga ega nazorat ostidagi o'quv quvuri yaratildi. Random Forest eng yaxshi natijalarni ko'rsatdi (MAPE: 13.35%, R^2: 0.922), chiziqli regressiya (MAPE: 15.66%) va ARIMA (MAPE: 27.10%) dan ustun keldi. Prognozlarni harakatlarga aylantirish uchun veb-boshqaruv paneli va aksiyalarni tavsiya qilish mexanizmi ishlab chiqildi va serverga joylashtirildi. Natijalar shuni ko'rsatadiki, ansambl usullari ko'p do'konli chakana savdo uchun prognozlash aniqligini sezilarli darajada yaxshilaydi, Korzinka uchun yillik ta'sir 150–200 milliard so'mni tashkil qiladi.

Абстракт. Эффективное управление запасами позволяет продуктовым магазинам предотвращать дефицит товаров, сокращать потери и более эффективно распределять ресурсы. В этом проекте рассматриваются проблемы прогнозирования, с которыми сталкивается компания «Корзинка». «Корзинка» — крупнейший в Узбекистане современный розничный продавец продуктов питания, имеющий более 190 магазинов трех форматов и обрабатывающий 94 миллиона транзакций в год. В годовом отчете компании за 2024 год указан уровень

доступности товаров в 92%, что ниже отраслевого стандарта в 95%. Уровень потерь в магазинах шаговой доступности составляет 4,5%. «Корзинка» также планирует увеличить свою сеть до 1000 магазинов к 2028 году. При таком масштабе ручного прогнозирования будет недостаточно. В рамках проекта было предложено возможное решение с помощью системы прогнозирования спроса на основе искусственного интеллекта путем внедрения и сравнения трех алгоритмов машинного обучения: линейной регрессии (базовый вариант), ARIMA (временные ряды) и случайного леса (ансамбль). Используя набор данных с Kaggle — «Спрос на товары в магазинах» (913 000 записей, 10 магазинов, 50 товаров, 5-летний период), был построен конвейер контролируемого обучения с 20 специально разработанными признаками. Наилучшие результаты показал алгоритм случайного леса (MAPE: 13,35%, R^2: 0,922), превзойдя линейную регрессию (MAPE: 15,66%) и ARIMA (MAPE: 27,10%). Была разработана и развернута на сервере веб-панель управления и система рекомендаций по акциям для преобразования прогнозов в действия. Результаты показывают, что ансамблевые методы значительно повышают точность прогнозирования для розничной торговли с несколькими магазинами, с предполагаемым ежегодным эффектом в 150–200 миллиардов сумов для Корзинки.

1. Introduction

1.1 Problem Overview

Accurately forecasting the demand for a product helps grocery retailers maintain their finances and customer satisfaction levels. If demand is overestimated it may lead to waste and later to a higher storage costs. Underestimating demand in the other hand, leads to stockouts, lost sales, and reduced customer trust. While the business is growing quickly and supply chains are less developed these problems become more challenging. For example, Uzbekistan's grocery market was worth USD 11.4 billion in 2023 and is expected to reach USD 17.7 billion by 2028 (Korzinka, 2024). Since modern retail accounts for only 9% of total grocery spending, there is significant room for growth and a strong need to improve operations as the market evolves.

1.2 Identifying the Problem

Korzinka's 2024 Annual Report was the main source for defining the main business problem. According to Korzinka's report, it has little domestic production and is double-landlocked (Korzinka, 2024, p. 29). Most modern grocery retailers target at least 95% availability; this 3% gap may result in approximately UZS 280 billion in lost sales annually. The Annual Report identifies improved forecasting as a key objective.

Shrinkage rates in the company's Mahalla stores reached 4.5% in 2024. While gross margins at 17.4% even 1% reduction in shrinkage might result in annual savings of nearly 94 billion soums. Fresh vegetable costs changed drastically in 2024 with carrots dropping 32% and onions falling 41%. The "First Price" strategy increased sales by 29% to 65% for key commodities. Accurate forecasting is crucial yet difficult due to unpredictable pricing, abrupt demand changes, and short product shelf life. In 2024, but this is still higher than the international standard of 18–22 days for similar products. The company's online channel, Korzinka Go, achieved 7 day turnover by improving forecasting. This shows that improved predictions can help optimize working capital (Korzinka, 2024, p.26).

The "Korzinka 3.0" expansion plan aims for about 1,000 stores by 2030, which is six times more than today. This growth will make forecasting much more complex, with store-SKU combinations increasing from 532,000 (152 stores × 3,500 SKUs) to 3.5 million. At this size, manual forecasting will not be possible.

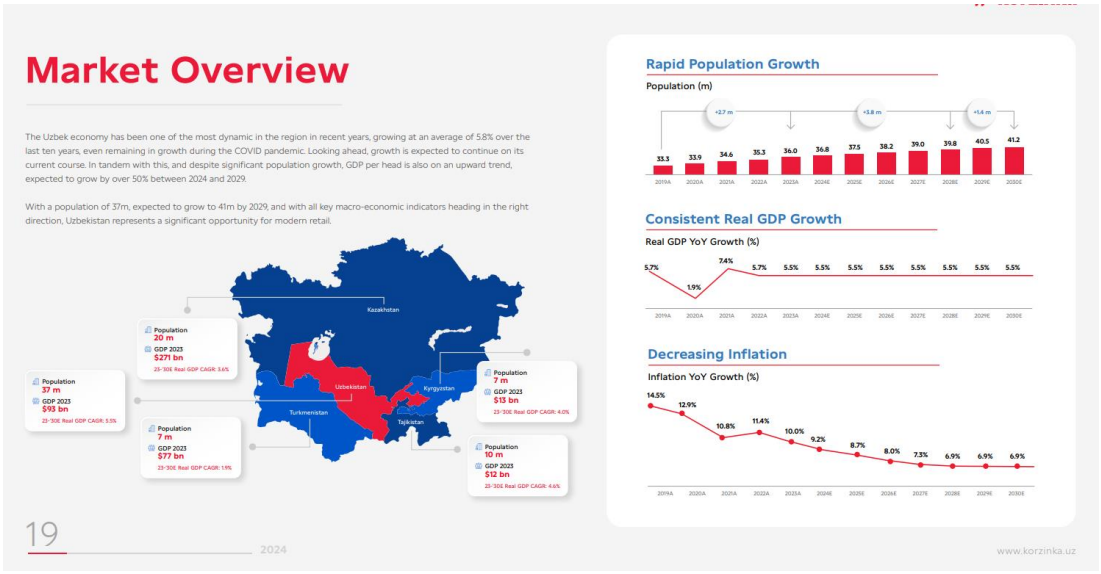


Figure 1 Market overview from Korzinka's report

1.3 Problem Statement

Korzinka does not have an advanced demand forecasting system that can accurately predict product-level demand across its multi-format store network. This limits the company's ability to reduce stockouts, minimise perishable waste, and manage the operational complexity of its planned six-fold expansion.

1.4 Objectives and Scope

Develop a demand forecasting system that achieves less than 15% Mean Absolute Percentage Error (MAPE) for 7-day sales predictions, in line with industry standards. Demonstrate the system's potential to improve product availability from 92% toward the 95% industry target and reduce shrinkage below 4.5%. Deliver actionable stock recommendations, including reorder points, safety stock, and order quantities, for supply chain planners.

Technical Objectives

To build a complete machine learning pipeline, build a pipeline that includes data ingestion, feature engineering (20 features), model training, and evaluation on a representative retail dataset. Develop a dashboard using Flask and React.js for forecast visualization and stock recommendation delivery. Dockerize the project and later deploy it to an Azure VM with 2 CPUs and 4 GB of RAM, implementing a CI/CD pipeline to speed up and simplify deployments. To guarantee repeatable, deployment-ready predictions, use joblib serialization alongside model persistence. proficiency with Python, scikit-learn, and statsmodels for developing whole machine learning pipelines. To become proficient in developing web applications by combining React frontends and Flask REST APIs with ML model inference.

Scope

MoSCoW prioritizing is used to make sure that the project's deliverables can be met within the academic time frame and solve the main business challenge.

Must Have: Need to compare three ML algorithms (Linear Regression, ARIMA, and Random Forest) on a dataset from a retailer with more than one store. Need to build a 20-feature engineering pipeline that has lag variables, rolling statistics, and calendar indications. After that, you need to utilize MAPE, RMSE, MAE, and R^2 to check the findings. Finally, need to develop a report that compares the models and looks at the mistakes for each one.

Should Have: Make a Flask REST API to show the best-performing model for inference; make a React.js dashboard to show the prediction; and make a stock recommendation engine to figure out reorder points, safety stock, and order amounts.

Could Have: Add 30 days to the forecast horizon in addition to 7 days; have supply chain professionals do a System Usability Scale (SUS) evaluation; set up a real-time stock tracking system

with a StockLevel table, POS integration, delivery receipt, and suggestions for transferring items between stores.

Will Not Have: Real-time streaming prediction pipeline; mobile application; deep learning models (LSTM, Transformer); deployment on Korzinka's production infrastructure; integration with SAP ERP.

2. Literature Review

2.1 An Overview of Demand Forecasting in Retail

Demand forecasting is an important part of supply chain management since it helps with choices about buying, restocking inventory, and planning logistics for distribution. Chopra and Meindl (2016, p.186) say that demand projections are the basis for all supply chain planning, such as manufacturing, transportation, and inventory management. In the grocery retail business, erroneous forecasting can have very bad effects. If you overestimate, you waste food and pay more to store it. If you underestimate, you run out of stock, lose sales, and lose customers' loyalty. Huber and Stuckenschmidt (2020) experimentally demonstrate that for perishable goods with high degradation rates, inaccurate demand predictions promptly result in overproduction (leading to waste) and empty shelves, thereby making daily restocking decisions heavily dependent on the accuracy of forecasts. Fildes, Ma, and Kolassa (2022), in their extensive review of retail forecasting research and practice, note that forecasts support allocation and replenishment decisions throughout the retail supply chain; however, no single technique attains absolute accuracy in all contexts—highlighting the necessity for domain-specific assessment of competing methodologies.

In recent years, using machine learning to predict demand has become more popular. A systematic review conducted by Douaioui et al. (2024), examining 119 papers from the Scopus database spanning 2015 to 2024, reveals an exponential increase in research utilizing machine learning and deep learning for supply chain forecasting, with nearly 73% of all publications emerging between 2021 and 2024. This trend shows that both academics and business are starting to realize that machine learning may greatly increase the accuracy of forecasts. According to McKinsey & Company (2017), AI-driven forecasting may cut supply chain mistakes by 20 to 50%, which can make products more available and operations more efficient. Nonetheless, the research warns that model performance is significantly context-dependent, fluctuating with data attributes, product categories, and forecasting horizons (Petropoulos et al., 2022), which justifies the comparison method employed in this project.

2.2 Linear Regression as a Baseline for Forecasting

Linear regression is still one of the most used baseline approaches used in demand forecasting research because it is easy to understand, efficient to apply, and easy to analyze (James et al., 2013). The approach uses conventional least squares minimization to find coefficients for a linear function that shows how one or more independent factors affect a dependent variable (in this example, sales amount). In retail forecasting, independent variables usually comprise time-based variables like the day of the week, the month, and seasonal indications, as well as lag factors that show recent sales trends.

The main benefit of linear regression is that it is easy to understand. Each coefficient explicitly measures how much its variable affects expected demand, which makes it easy for those who aren't technical, like supply chain planners, to use. But the linearity assumption is often not true in stores because demand is affected by non-linear interactions between promotions, seasonality, and outside influences (James et al., 2013). Schonlau and Zou (2020) show that decision tree ensembles like Random Forest are better at capturing these non-linearities and usually provide more accurate predictions than standard least squares regression. Sharma et al. (2022) validated the efficacy of linear regression as a reliable benchmark inside hybrid ensemble frameworks, employing it in conjunction with Random Forest and XGBoost in a retail sales forecasting analysis. So, linear regression is used in this research as a baseline to set a performance floor against which more advanced methods can be tested.

2.3 ARIMA Models for Predicting Time Series

The Autoregressive Integrated Moving Average (ARIMA) model, formalized by Box and Jenkins (1976), is a well-established statistical method for time series forecasting. ARIMA models temporal relationships using three parts: autoregression (AR), which looks at the link between an observation and lagged observations; integration (I), which uses differencing to make the data stationary; and moving average (MA), which looks at how residual errors rely on each other. People usually choose the parameters (p, d, q) using information criteria like the Akaike Information Criterion (AIC) or automated selection methods like auto_arima (Hyndman and Athanasopoulos, 2018).

ARIMA and its seasonal extension (SARIMA) have exhibited robust efficacy in areas marked by distinct trend and seasonal patterns (Hyndman and Athanasopoulos, 2018). ARIMA has been utilized in retail forecasting for product categories that have consistent cyclical demand, exemplified as fresh produce with weekly purchasing trends.

However, the research delineates several substantial constraints of ARIMA within the retail domain. First, ARIMA is primarily a univariate approach; it cannot inherently integrate exogenous factors such as store features, promotional activities, or weather data without modifications like ARIMAX (Makridakis, Spiliotis, and Assimakopoulos, 2018). Second, the assumption of linearity makes it harder for it to pick up on the complicated, non-linear demand patterns that are common in multi-store retail operations. Third, ARIMA needs a different model for each time series. In a retail setting with 10 stores and 50 goods, this implies fitting 500 separate models, which is hard to scale up to Korzinka's operational level of over 500,000 store-SKU combinations. Falatouri et al. (2023) and Huber and Stuckenschmidt (2020) both say that ensemble machine learning approaches always work better than ARIMA on retail datasets. This is especially true when the data has complicated multi-variate correlations that ARIMA's univariate structure can't use. Falatouri et al. (2023) discovered that this performance disparity is particularly evident in perishable product categories, since ARIMA's failure to integrate external influences, like weather and promotional events, severely restricts its predicting capacity.

2.4 Random Forest: Using Ensemble Learning to Predict Demand

Breiman (2001) came up with Random Forest, which is an ensemble learning approach that builds a lot of decision trees during training and gives the average prediction over all trees for regression problems. Each tree is trained on a bootstrap sample of the training data (bagging), and at each split, only a random subset of characteristics is used. This makes trees different from each other and stops them from overfitting. Breiman (2001) demonstrated that the generalization error of a random forest diminishes as the quantity of trees rises, rendering the approach robust against overfitting, even with substantial numbers of estimators. The built-in feature significance measure, which shows how much each variable helps reduce variation across all trees, is very useful for understanding how to use the model in real life (Schonlau and Zou, 2020).

In the literature on retail forecasting, Random Forest has become one of the most reliable models. Falatouri et al. (2023) evaluated tree-based ensemble models against Long Short-Term Memory (LSTM) deep learning networks using over six years of retail demand data from perishable product categories, concluding that tree-based ensembles surpassed LSTM in all evaluation metrics, including MAPE, MAE, RMSE, and R^2. The difference in performance was especially clear for fresh meat items, where ensemble models cut MAPE by more than 25% compared to LSTM. This is similar to Korzinka's high-shrinkage perishable lines. Huber and Stuckenschmidt (2020) also showed that machine learning methods, like gradient-boosted decision trees, worked much better than traditional statistical methods for predicting daily demand in a bakery chain and were better for use in many stores at once.

The main benefit of Random Forest over ARIMA for this project is that it can handle more than one variable at a time. Random Forest may include temporal information (including the day of the week, month, and lag variables), store-level characteristics, rolling statistics, and other contextual factors in a single model. This is not possible with ARIMA. This means that the algorithm may learn patterns that are common to all the stores and items in the dataset instead of having to fit different models for each store-item combination. But Random Forest does have some problems. For example,

it can't generate predictions outside of the range of the training data (the minimum and maximum values of the training set restrict the forecasts). Also, because it is an ensemble, individual predictions are less clear than a single ARIMA equation (James et al., 2013).

2.5 Feature Engineering for Retail Time Series

The efficacy of machine learning models in demand forecasting is fundamentally contingent upon the quality and pertinence of designed characteristics. Feature engineering turns raw time series data into a supervised learning issue by making useful input variables that show temporal patterns, recent trends, and context.

Lag characteristics, which show the sales value a certain number of time steps ago (for example, 7 days ago, 14 days ago, or 30 days ago), are often seen as the most crucial factors in retail demand models. Falatouri et al. (2023) discovered that lag characteristics and rolling statistics constituted the predominant sources of predictive power in their tree-based ensemble models. Rolling window statistics, such the 7-day rolling mean and standard deviation, show short-term momentum and volatility. This gives the model a better picture of the recent demand trend than just one point in time. Calendar characteristics like day of the week, month, and weekend indications highlight patterns of demand that happen over and over again. External factors like holiday calendars and weather data have also been proved to make forecasts for perishable categories even more accurate. Huber and Stuckenschmidt (2020) show that adding special days to machine learning models greatly improved daily demand projections. This is because holidays and promotional events produce demand patterns that historical lag characteristics alone can't predict.

In Uzbekistan, extrinsic elements that are especially important are the times of Ramadan and Nowruz, very hot or very cold weather (−15°C to 50°C), and the times of year when crops are harvested. Falatouri et al. (2023) included weather and holiday data in their European retail analysis; however, no published research has particularly investigated these elements within a Central Asian retail context—an identified gap that this project recognizes as a constraint and a pathway for future research.

2.6 Metrics for Evaluating Forecasting Models

Choosing the right assessment criteria is important for making meaningful comparisons between models. Hyndman and Koehler (2006) offer the conclusive examination of forecast accuracy metrics, emphasizing the advantages and drawbacks of each. The Mean Absolute Percentage Error (MAPE) is the most common measure used in retail forecasting since it can be used to compare items and outlets with various demand volumes directly. But MAPE doesn't work when the actual values are zero, and it might be asymmetrically biased, punishing under-predictions more than over-predictions. Root Mean Squared Error (RMSE) is a scale-dependent measure that is sensitive to big mistakes. This makes it useful for finding models that sometimes make very bad predictions, which is very important in managing perishable inventory because one big over-order can cause a lot of waste. Mean Absolute Error (MAE) is a better choice since it is less susceptible to outliers. The coefficient of determination (R^2) shows how much of the variation in demand the model explains, which is a good way to estimate how well the model fits. This study uses all four metrics—MAPE, RMSE, MAE, and R^2—to give a full, multi-dimensional assessment, following the methods suggested by Hyndman and Koehler (2006) and used in many recent studies on retail forecasting (Falatouri et al., 2023; Sharma et al., 2022).

2.7 Recognized Knowledge Deficiency

The literature on machine learning for retail demand forecasting is substantial and quickly expanding; yet, significant gaps exist that this study seeks to solve. First, most of the studies that have been published look at mature markets in North America, Europe, and East Asia. There is a clear lack of research on demand forecasting in Central Asian retail markets, where supply chain conditions—such as double-landlocked geography, extreme climate variation, and low modern retail penetration (9% in Uzbekistan)—present unique challenges. Petropoulos et al. (2022) assert that predicting effectiveness is significantly influenced by context; yet, no empirical research have evaluated ML forecasting systems in these particular circumstances.

Second, whereas several research contrast two methodologies (usually a statistical baseline versus a single machine learning model), fewer offer a comprehensive three-way comparison of linear regression, ARIMA, and Random Forest applied to the identical dataset with uniform assessment criteria. Third, most of the time, the literature just talks about how accurate models are and not on how to turn forecasts into useful supply chain choices like reorder points and safety stock estimates. This project helps close these gaps by testing three very different ways of making predictions in a model that is similar to Korzinka's operational environment. It also adds a decision-support dashboard to the system that gives supply chain planners useful stock recommendations.

3. Research Framework

The project employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach (Chapman et al., 2000). We picked CRISP-DM over SEMMA and KDD because it starts with understanding the business, maintains technical work in line with the aims of the organization, is iterative to allow for continual re-evaluation, and is tool-agnostic and well recognized. The project's structure is based on its six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

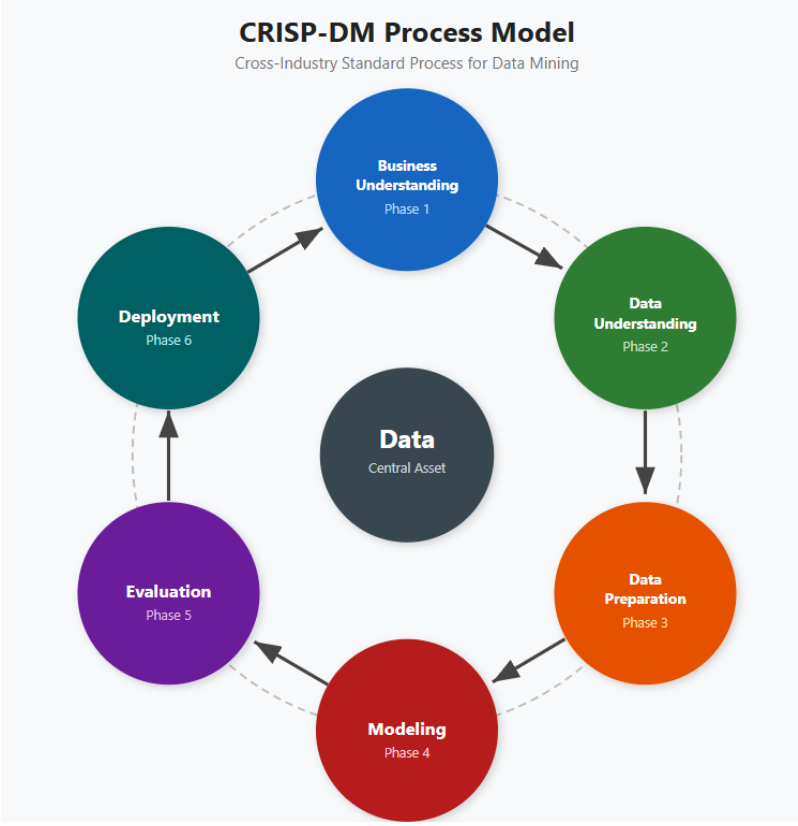


Figure 2 CRISP-DM cycle diagram

3.1 Market Analysis

The food retail sector in Uzbekistan is worth $11.4 billion (2023) and is predicted to expand to $17.7 billion by 2030. This is due to a 5.5% annual GDP growth rate, a population expansion from 37 million to 41 million by 2029, and a drop in inflation from 10% (2023) to 6.9% (2030) (Korzinka, 2024). Only 9% of grocery expenditure goes to modern stores; traditional bazaars are still the most popular places to shop. Retailers who invest in technology and operational efficiency have a lot of chances because this industry isn't very big yet.

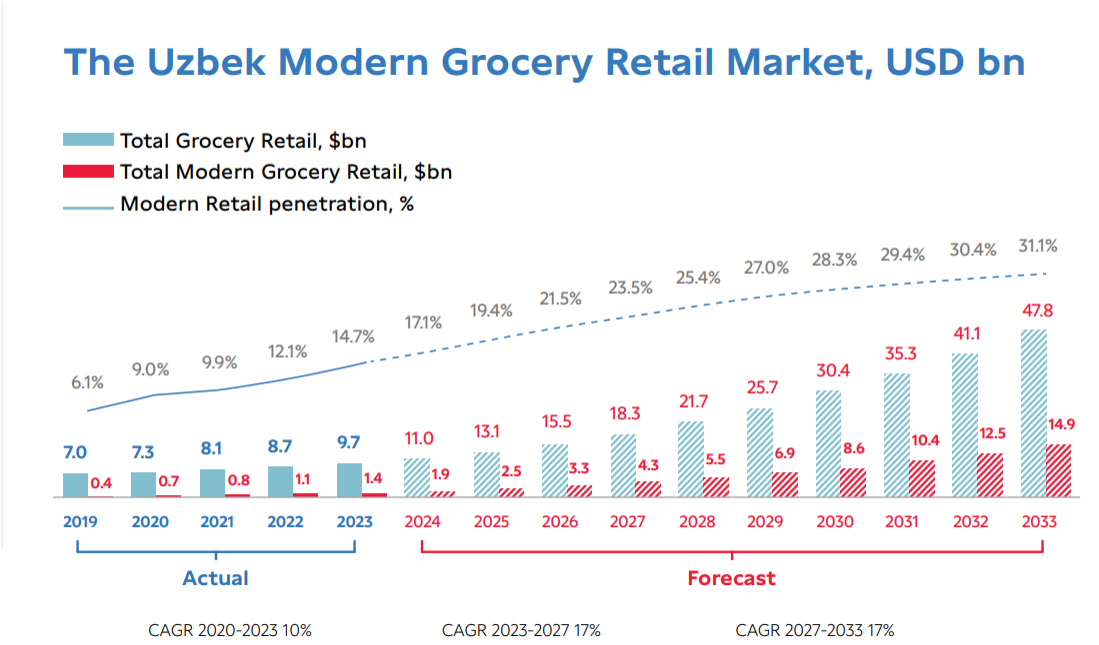


Figure 3 Uzbekistan grocery market size chart (2023–2033 projection)

Korzinka leads the market with brand awareness six times higher than its nearest competitor, a Net Promoter Score of 83, annual sales of UZS 9.4 trillion, and 250,000 daily store visitors. Its loyalty programme, with 1.8 million active cardholders accounting for 55% of transactions, offers a strong data foundation for demand modelling.

3.3 Competitor Analysis

Competitors are judged on how well they can make predictions and how big their operations are to show how useful an ML-based system is.

Table 1Uzbekistan Grocery Retail Competitor Landscape

Retailer	Stores	Format
Korzinka	152	Supermarket, proximity, online
Havas	340+	Discount, limited fresh
Makro	56	Relaunching formats

There is no evidence in publicly available information that Korzinka's main competitors have invested in machine learning-based forecasting systems. Uzbekistan's modern retail market is still in its early stages, with only 9% of the population using it. This means that advanced forecasting abilities could give businesses an edge over their competitors, especially as the market becomes more formalized and operational efficiency becomes more important for competitive positioning.

3.4 SWOT Analysis

Table 2 SWOT Analysis of the Forecasting System

Strengths	Weaknesses
Random Forest achieves 13.35% MAPE, exceeding the 15% target. Single unified model scales to all store-item combinations. Feature importance provides explainable predictions for	Built on proxy dataset, not Korzinka's actual data. Lacks external variables (weather, holidays, promotions). ARIMA evaluated on

planners. Stock recommendation engine bridges prediction to action.	subset only. No formal usability testing with real planners conducted.
Opportunities	Threats
New DC (Q3 2025) requires intelligent stock allocation. Korzinka 3.0 expansion to 1,000 stores demands automated forecasting. Board-level commitment to analytics investment. 1.8M loyalty members provide rich data.	Proprietary data access may be restricted. Competitors may adopt similar technology. SAP ERP integration complexity. Resistance to change from manual processes. Uzbekistan-specific demand patterns may differ from proxy data.

3.5 Analysis of Users

There are three primary types of users of the forecasting system at Korzinka, each with their own needs:

Store Manager (like the Mahalla proximity store): Makes sure that 3,500 to 3,700 SKUs are restocked every day. Instead of raw projections, it needs concise, actionable outputs like reorder amounts and warnings for high-risk goods. The fundamental problem is that too much fresh food goes to waste because of over-ordering, which is made worse by a 4.5% shrinkage rate. A dashboard should show which things to order and how much, with perishability risk as the main factor.

Central Supply Chain Planner: Makes sure that merchandise is distributed evenly throughout 152 retailers and works with the new distribution center. Needs overall projections by store cluster, category, and time frame for planning logistics. The biggest problem is that when there are stockouts, emergency restocks have to happen, which means incomplete truck loads and inefficient delivery routes that are now averaging 160 kilometers per shop. Needs to look at trends and report exceptions.

The Category Manager is in charge of planning the selection and promotion of all product categories, including 511 private label items (7.6% of sales, with a goal of 20–25% by 2027). Needs a sensitivity study of demand to see how projections vary as prices or sales change. The biggest problem is that there isn't much historical data for new private label launches.

3.6 Reasons for the Dataset

Because Korzinka's transactional data is susceptible to business, the Kaggle Store Item Demand Forecasting dataset was chosen as a structurally comparable substitute. It has daily sales records for 10 locations and 50 goods from 2013 to 2017, for a total of 913,000 records with no missing values. The dataset has a multi-store structure, item-level information, seasonal trends, and demand that changes over time, all of which are similar to Korzinka's environment. The 500 store-item combinations are a good sample of how hard it is to make predictions at Korzinka's level

4. Computing Methodology

The first part of the study looked at four possible algorithms: ARIMA, Random Forest, and XGBoost. After doing some initial research, XGBoost was left out so that there would be a clear three-way comparison between three very distinct paradigms: a linear baseline, a classical time series technique, and an ensemble learning approach. During development, it became clear that ARIMA had structural limits, which confirmed that the three models chosen were different enough to allow for relevant conclusions about how to choose a forecasting approach.

Analyzed both Django and Flask to help us choose the backend framework. Flask was chosen because its lightweight design is ideal for a focused ML inference API with 9 endpoints. Django's monolithic structure would make things more complicated than they need to be for this project. React.js for the front end instead of simpler options like vanilla JavaScript or jQuery so that we could create a responsive single-page application with interactive charts. This is important for a dashboard that supply chain planners would use every day.

4.1 Technology Stack

Table 3 Technology Stack and Justification

Layer	Technology	Justification
Data / ML	Python 3.9+, pandas, NumPy, scikit-learn, statsmodels, pmdarima	Dominant data science ecosystem. scikit-learn provides unified API for model training, evaluation, and persistence. pmdarima offers automatic ARIMA parameter selection via AIC.
Analysis	Jupyter Notebooks, Matplotlib, Seaborn	Iterative documented analysis combining code, output, and narrative. Publication-quality visualisations for EDA and diagnostics.
Backend	Flask, PostgreSQL	Lightweight Python REST API framework for ML model inference endpoints. PostgreSQL for relational storage of sales data and prediction logs.
Frontend	React.js 18, Recharts, Vite	Component-based SPA for interactive forecasting dashboard. Recharts for trend visualisation, model comparison displays, and stock recommendation charts. Vite provides fast development builds.
DevOps	Git, GitHub, VS Code, joblib	Version control with branching for model experiments. Model serialisation to .pkl files with fixed random seeds (random_state=42) ensuring reproducibility.
Deployment	Azure App Service (2 vCPU, 4 GB RAM)	Cloud-hosted deployment for the Flask backend and React frontend. Provides managed scaling, HTTPS, and continuous deployment from GitHub.

4.2 The Structure of the System

The system is built on a four-layer modular pipeline design. The data layer takes in and checks raw CSV files. The feature engineering layer turns the raw dataset with four columns (date, store, item, sales) into a supervised learning matrix with 20 features. These features include calendar features (year, month, day, day of week, day of year, week, quarter, weekend indicator), lag features (7-day, 14-day, 30-day sales lags), and rolling statistics (7-day, 14-day, 30-day rolling means and standard deviations). The first 30 days are taken out since there isn't enough lag history, therefore the dataset is now just 898,000 entries. The model layer trains three algorithms using a temporal validation split of 852,000 training records and 46,000 validation records. The deployment layer shows out the best model using Flask REST endpoints that the React dashboard uses.

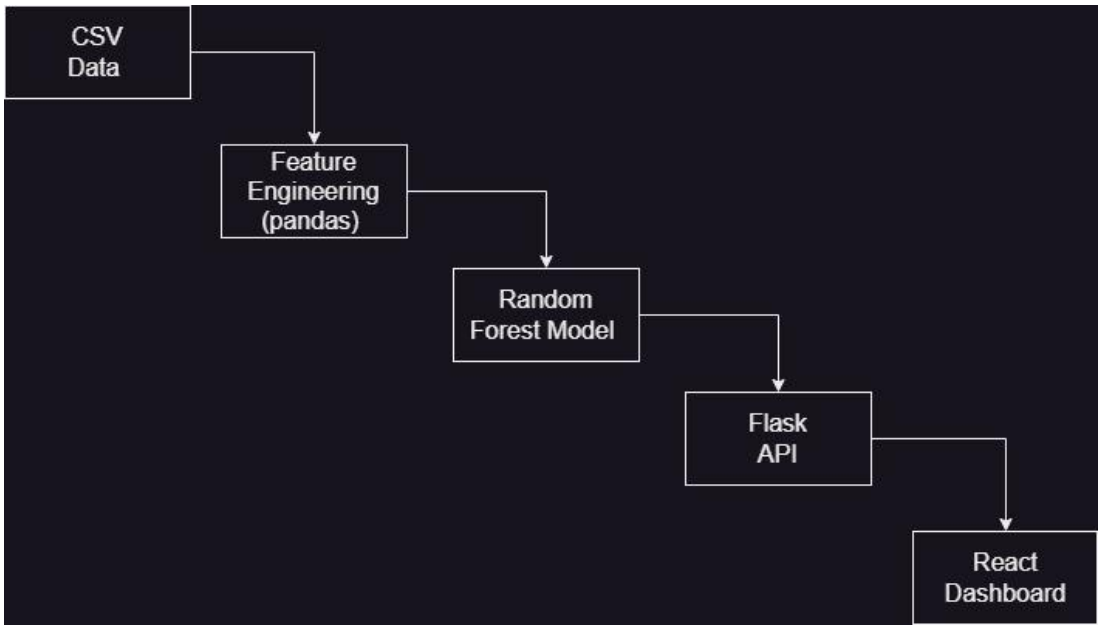


Figure 4 System architecture diagram

4.3 Design Documents

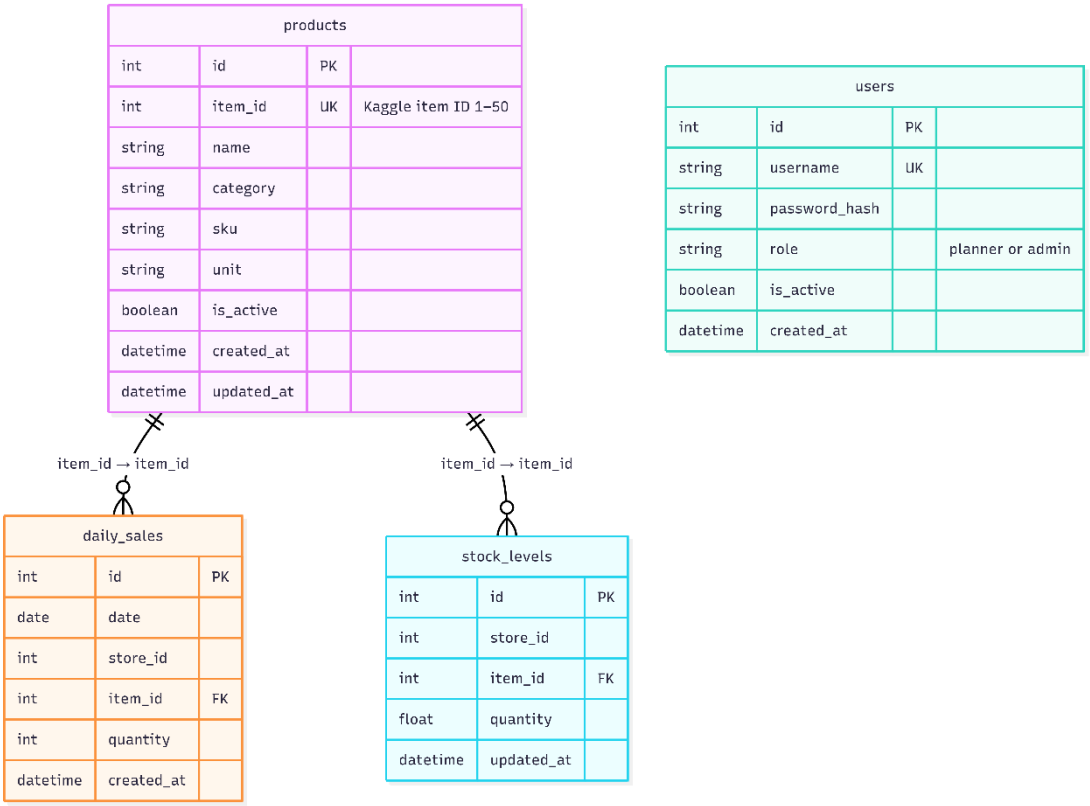


Figure 5 ER diagram for database

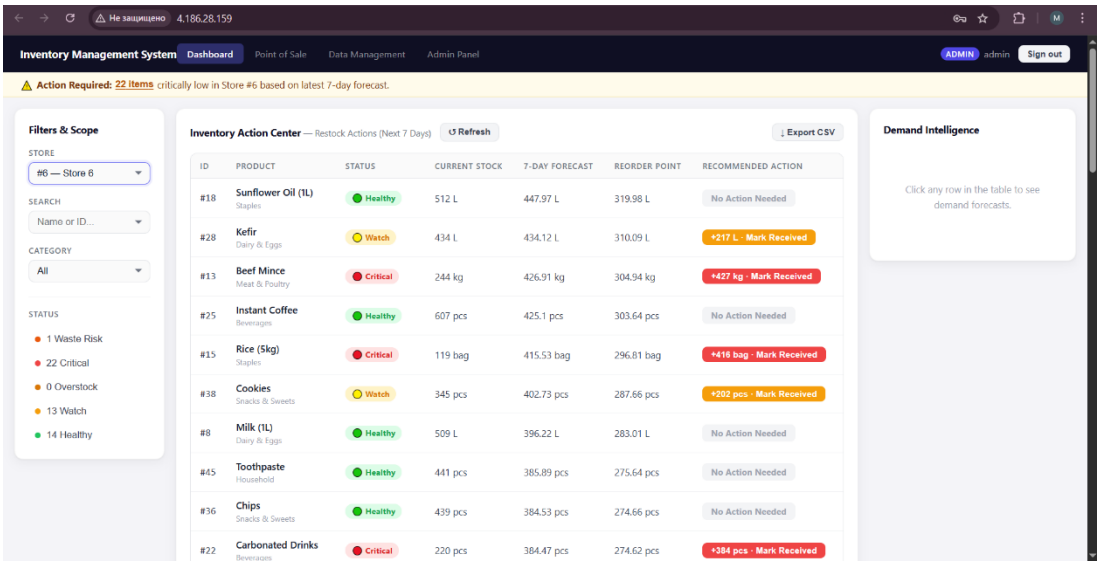


Figure 6 Dashboard UI

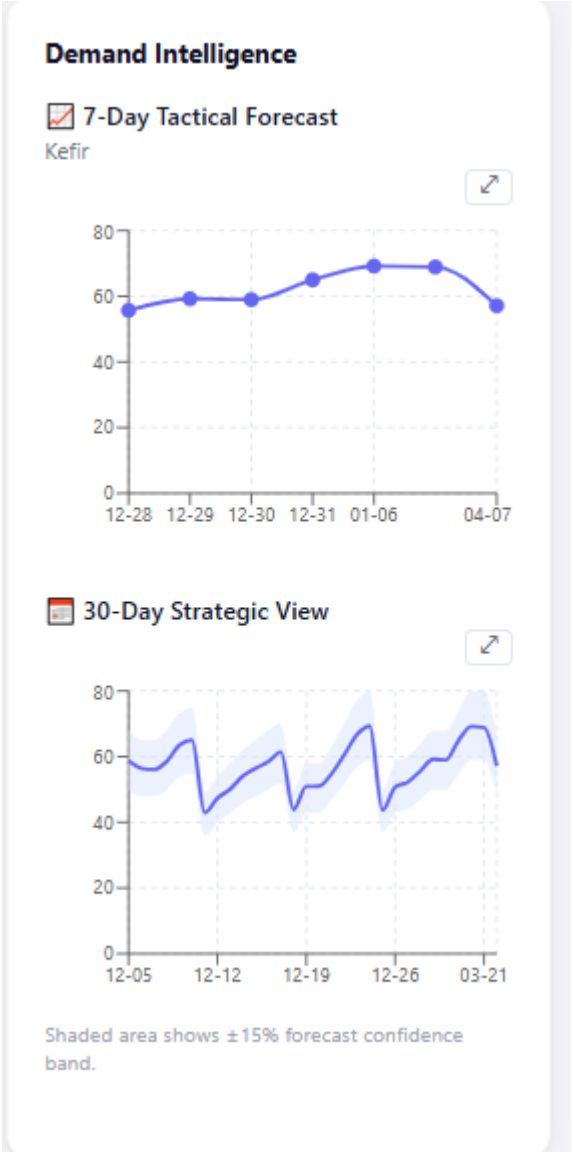


Figure 7 Dashboards for 7-day and 30-day forecast

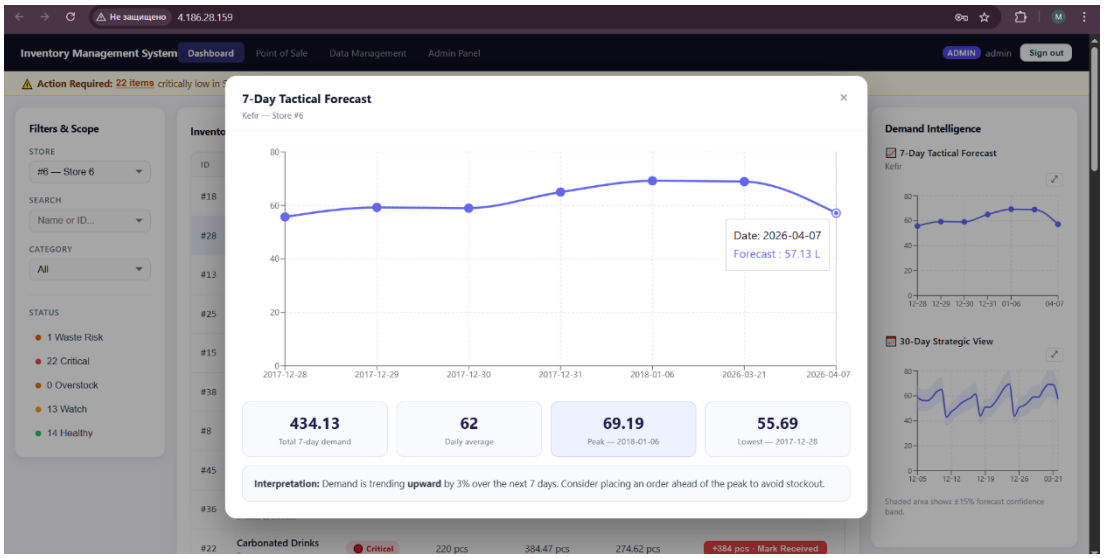


Figure 8 Full screen Dashboard view for 7 days.

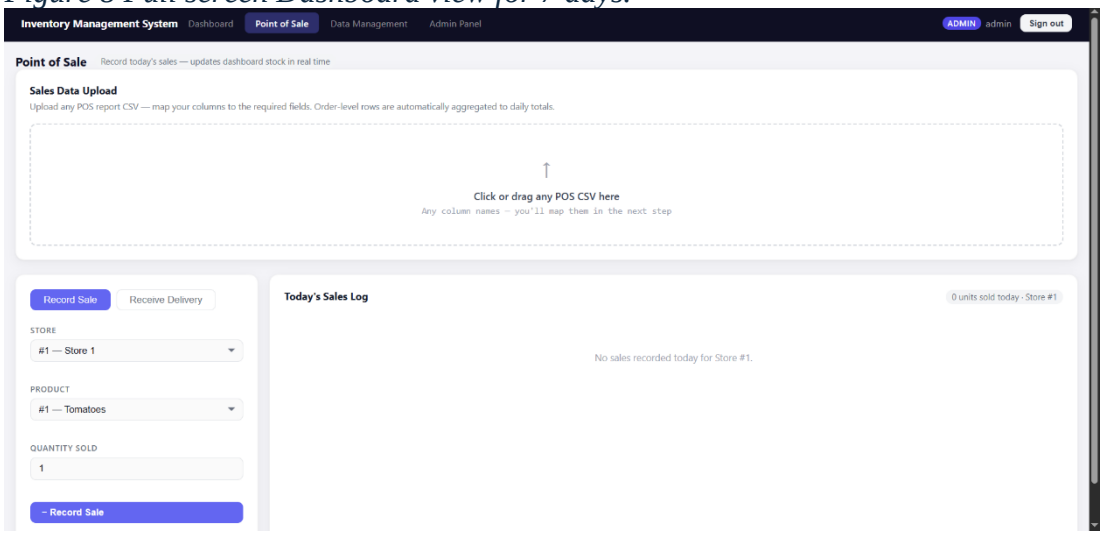


Figure 9 POS view

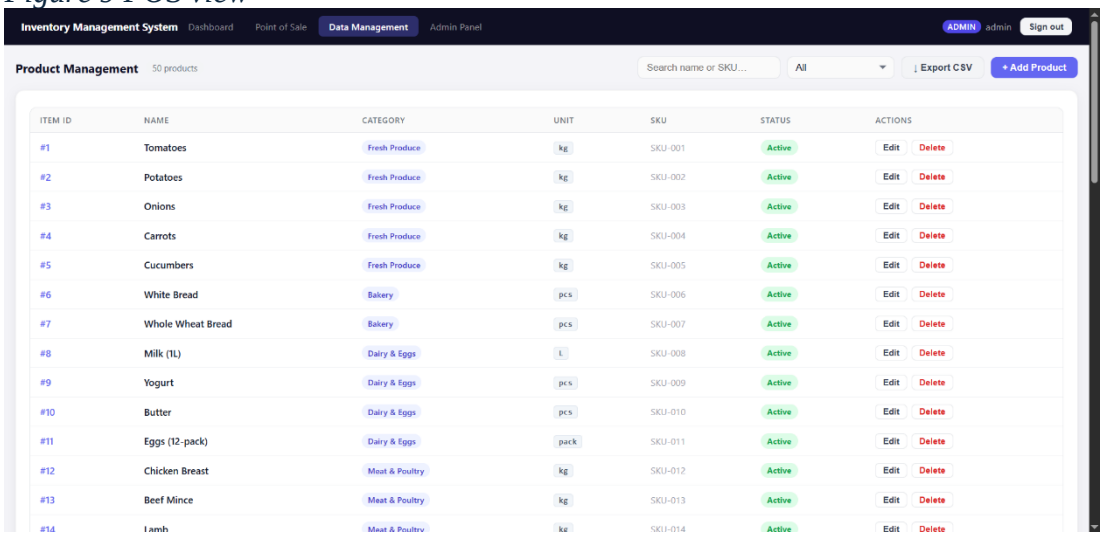


Figure 10 Data management panel.

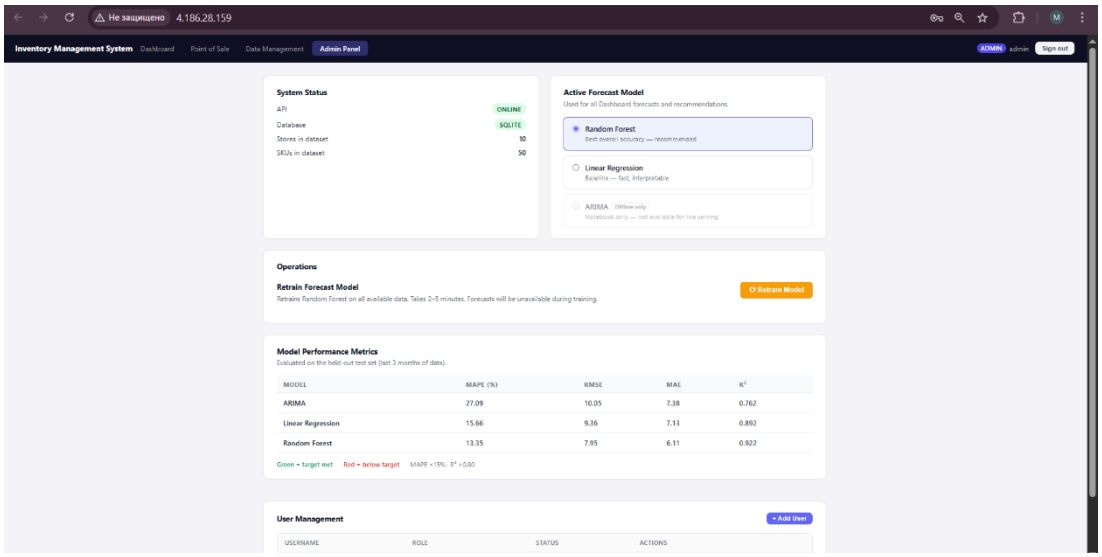


Figure 11 Admin panel

4.5 Testing

Testing was conducted at three levels, documented in the table below.

Table 4 Testing Summary

Test Case	Method	Status
Zero missing values in raw data	pandas isnull().sum()	Pass
Date range 2013–2017 confirmed	min/max date assertion	Pass
913,000 records loaded correctly	Shape validation	Pass
20 features engineered	Column count check	Pass
No data leakage in temporal split	Max train date < min val date	Pass
LR predictions non-negative	np.clip validation	Pass
RF MAPE < 15% target	Metric evaluation	Pass (13.35%)
ARIMA residual diagnostics	Ljung-Box test	Fail ($p < 0.001$)
Stock recommendations logical	Non-negative, reorder < forecast	Pass
Flask API forecast endpoint	POST request with store/item	Pass
Dashboard renders forecasts	Browser functional test	Pass

5. Results

5.1 Model Comparison

Random Forest did the best on all measures, with a MAPE of 13.35%, which was much below the project's 15% aim. The model accounted for 92.2% of the variation in demand ($R^2 = 0.922$) and

had a mean absolute error of about 6 units per day estimate. Linear Regression got a MAPE of 15.66%, which was just above the objective. ARIMA, on the other hand, did far worse with a MAPE of 27.10%.

Table 5 Model Comparison Results (Validation Set: Oct–Dec 2017, 46,000 records)

Model	MAPE (%)	RMSE	MAE	R^2
Random Forest	13.35	7.948	6.105	0.922
Linear Regression	15.66	9.358	7.126	0.892
ARIMA	27.10	10.054	7.376	0.762



Figure 12 MAPE/RMSE/MAE across 3 models

5.2 Per-Item Analysis

Random Forest did better than Linear Regression on all 50 items, with an average improvement of 2.32 percentage points in MAPE. Item 5 (5.45 pp), which was the most variable (CV = 0.396), had the most improvement. 34 of the 50 items hit the 15% MAPE goal using Random Forest, whereas only 29 did with Linear Regression. The median absolute error was 4.9 units for RF and 5.6 units for LR, which shows that RF was better over the whole error range.

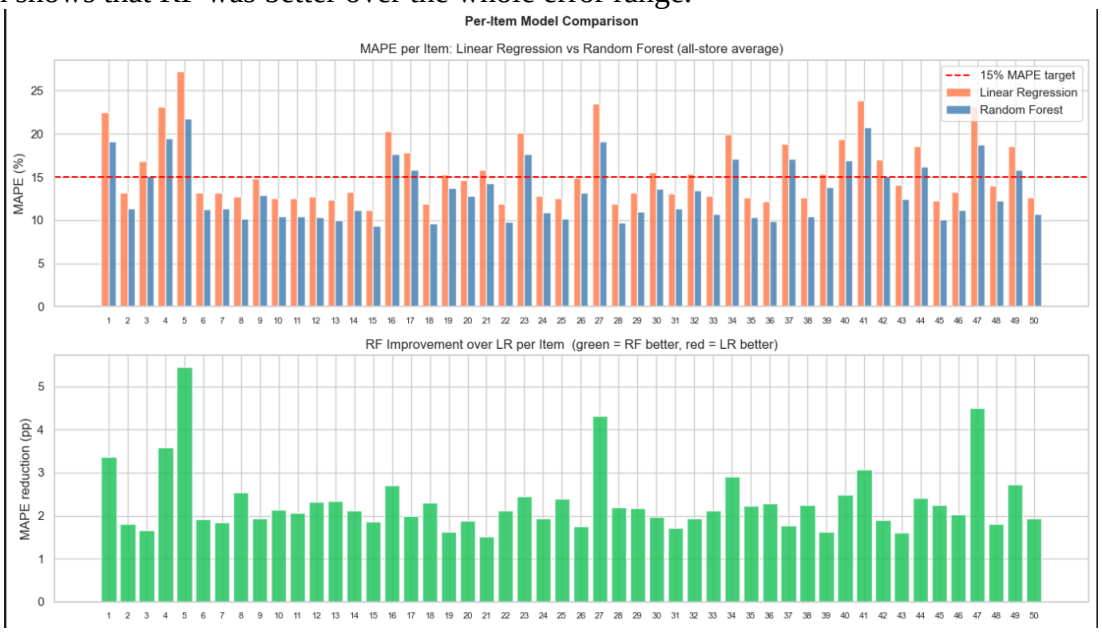


Figure 13 Per-item MAPE comparison chart — RF vs LR across all 50 items

5.3 Feature Importance

The 7-day sales lag (lag_7) was the most important predictor, making up 47.0% of the overall variance reduction. The 7-day rolling mean added 23.6% and the 14-day rolling mean added 15.8%. Together, these three characteristics made up 86.4% of the predictive value. Calendar features added a small amount (day_of_week at 2.2%), which confirmed that there is a weekly pattern. Store and item

IDs added less than 0.1% each, which suggests that demand patterns are mostly time-based at this level of aggregation.

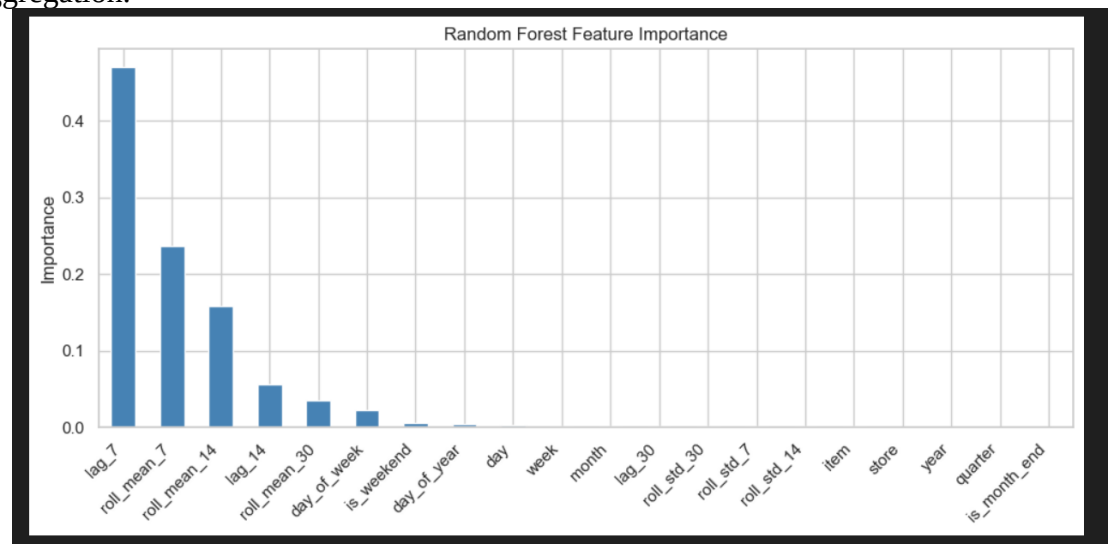


Figure 14 Feature importance bar chart from Random Forest

5.4 ARIMA Diagnostics

Residual analysis was used to find out why ARIMA didn't work well. The Ljung-Box test showed strong autocorrelation at lags 14 and 30 ($p < 0.001$), which showed that the model had not properly captured temporal relationships. Three fundamental reasons were identified: the univariate structure cannot leverage cross-item and cross-store trends; the linearity assumption obstructs the modeling of non-linear feature interactions; and fitting 500 distinct models leads to parameter estimation instability for products with erratic demand.

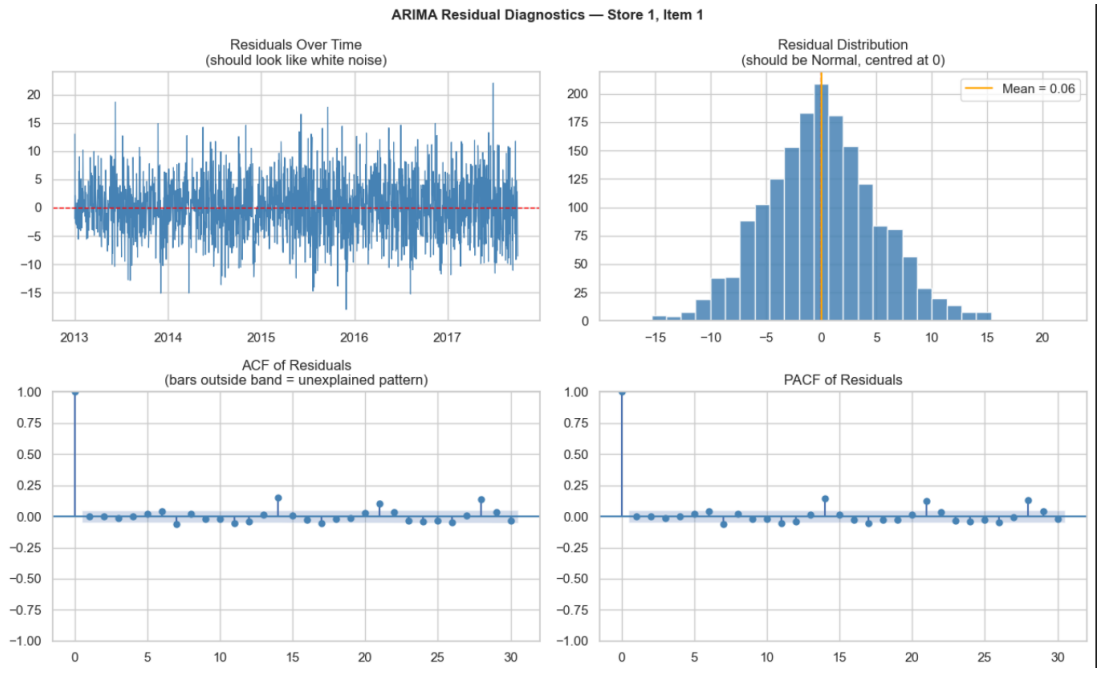


Figure 15 ARIMA residual diagnostics

5.5 Stock Recommendations

The stock suggestion engine created reorder points, safety stock (a 3-day buffer), and order amounts for all 500 combinations of stores and items. The pair with the most demand (store 2, item 28) had a 7-day forecast of 601 units, a reorder point of 429 units, and a safety stock of 258 units.

Top 10 highest-demand store-item pairs:

store	item	forecast_7d	avg_daily	safety_stock	reorder_point	order_qty
2	28	601.166869	85.880981	257.6	429.4	601
8	15	597.957041	85.422434	256.3	427.1	598
2	15	586.663092	83.809013	251.4	419.0	587
2	25	585.807757	83.686822	251.1	418.5	586
2	45	585.616183	83.659455	251.0	418.3	586
2	38	585.155547	83.593650	250.8	418.0	585
8	13	584.210765	83.458681	250.4	417.3	584
8	38	580.396325	82.913761	248.7	414.5	580
8	28	579.210080	82.744297	248.2	413.7	579
2	13	577.887635	82.555376	247.7	412.8	578

Figure 16 Stock recommendation output table for top 10 store-item pairs

6. Conclusions

The results show that ensemble machine learning methods are far more accurate than both traditional statistical methods and simple regression baselines for predicting demand in multi-store retail. The 13.35% MAPE for Random Forest is 49% better than ARIMA and 15% better than Linear Regression. This is in line with what Falatouri et al. (2023) and Huber et al. (2019) found.

The 7-day lag feature is the most important (47%), which has direct effects on Korzinka. The most recent sales history is the best predictor of near-term demand, which supports the idea of integrating real-time point-of-sale data into any manufacturing system. The shortcoming of ARIMA supports the consensus in the literature that univariate approaches are not enough for retail complexity that is multi-dimensional and large-scale.

In Korzinka's case, a 13% MAPE model might help raise the availability from 92% to 95% by making replenishment more precise. The stock suggestion engine makes it possible for forecasts and choices to work together by creating reorder points that balance the danger of running out of stock with the risk of shrinking. The expected growth to 3.5 million store-SKU combinations needs a single Random Forest model that can grow with it, rather than 500 individual ARIMA models.

7. Reflection and Future Work

7.1 Limitations

The biggest problem is that Korzinka's real data wasn't used; instead, a proxy dataset was used. The Kaggle dataset does not include promotional calendars, weather data, cultural holidays (such as Navruz and Ramadan), product prices, or the difference between perishable and non-perishable items. The 10-store scale doesn't really assess performance at 532,000 choices. We could only test ARIMA on a 2-store, 5-item subset because of limits on computing power. No official System Usability Scale (SUS) test was done with genuine supply chain planners.

7.2 Future Development

The system must be validated using Korzinka's real data, which includes factors peculiar to Uzbekistan. Should be tested more complex algorithms like XGBoost, LightGBM, and LSTM. Anomaly alerts, SAP ERP connection, and role-based access should all be added to the dashboard. For planning at the distribution center, the forecasting horizon should be 30 days long. A rigorous SUS evaluation with people who work in the supply chain would prove that it is useful in real life.

References

1. Box, G.E.P. and Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.
2. Breiman, L. (2001) 'Random Forests', Machine Learning, 45(1), pp. 5–32. Available from https://doi.org/10.1023/A:1010933404324 [Accessed 15 January 2026].
3. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000) CRISP-DM 1.0: Step-by-Step Data Mining Guide. SPSS Inc.
4. Chopra, S. and Meindl, P. (2016) Supply Chain Management: Strategy, Planning, and Operation. 6th edn. Harlow: Pearson Education.

5. Douaioui, K., Oucheikh, R., Benmoussa, O. and Mabrouki, C. (2024) 'Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review', Applied System Innovation, 7(5), p. 93. Available from https://doi.org/10.3390/asi7050093 [Accessed 20 February 2026].
6. Falatouri, T., Brandtner, P., Udokwu, C. and Darbanian, F. (2023) 'Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning', Applied Sciences, 13(19), p. 11112. Available from https://doi.org/10.3390/app131911112 [Accessed 10 January 2026].
7. Fildes, R., Ma, S. and Kolassa, S. (2022) 'Retail Forecasting: Research and Practice', International Journal of Forecasting, 38(4), pp. 1283–1318. Available from https://doi.org/10.1016/j.ijforecast.2019.06.004 [Accessed 5 February 2026].
8. Huber, J. and Stuckenschmidt, H. (2020) 'Daily Retail Demand Forecasting Using Machine Learning with Emphasis on Calendric Special Days', International Journal of Forecasting, 36(4), pp. 1420–1438. Available from https://doi.org/10.1016/j.ijforecast.2020.02.005 [Accessed 12 January 2026].
9. Hyndman, R.J. and Athanasopoulos, G. (2018) Forecasting: Principles and Practice. 2nd edn. Melbourne: OTexts. Available from https://otexts.com/fpp2/ [Accessed 8 January 2026].
10. Hyndman, R.J. and Koehler, A.B. (2006) 'Another Look at Measures of Forecast Accuracy', International Journal of Forecasting, 22(4), pp. 679–688. Available from https://doi.org/10.1016/j.ijforecast.2006.03.001 [Accessed 15 January 2026].
11. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning: With Applications in R. New York: Springer. Available from https://doi.org/10.1007/978-1-4614-7138-7 [Accessed 10 January 2026].
12. Korzinka (2024) Annual Report 2024. Tashkent: Anglesey Food FE LLC. Available from https://korzinka.uz/ [Accessed 12 February 2025].
13. Korzinka (2025) 'Korzinka Secures $110 Million Equity Investment to Accelerate Store Expansion and Advance Modern Retail in Uzbekistan', Pivot, 1 April. Available from https://pivot.uz/korzinka-secures-110-million-equity-investment-to-accelerate-store-expansion-and-advance-modern-retail-in-uzbekistan/ [Accessed 5 April 2026].
14. Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018) 'Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward', PLoS ONE, 13(3), e0194889. Available from https://doi.org/10.1371/journal.pone.0194889 [Accessed 20 January 2026].
15. McKinsey & Company (2017) Smartening Up with Artificial Intelligence (AI): What's in It for Germany and Its Industrial Sector? McKinsey Global Institute. Available from https://www.mckinsey.com/capabilities/quantumblack/our-insights/smartening-up-with-artificial-intelligence [Accessed 25 January 2026].
16. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Ben Taieb, S., Bergmeir, C., Bessa, R.J., Bijak, J., Boylan, J.E. et al. (2022) 'Forecasting: Theory and Practice', International Journal of Forecasting, 38(3), pp. 705–871. Available from https://doi.org/10.1016/j.ijforecast.2021.11.001 [Accessed 15 February 2026].
17. Schonlau, M. and Zou, R.Y. (2020) 'The Random Forest Algorithm for Statistical Learning', The Stata Journal, 20(1), pp. 3–29. Available from https://doi.org/10.1177/1536867X20909688 [Accessed 18 January 2026].
18. Sharma, A., Raut, R.D., Mangla, S.K., Narkhede, B.E., Luthra, S. and Gokhale, R. (2022) 'A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach', Operations Research Forum, 3(4), p. 166. Available from https://doi.org/10.1007/s43069-022-00166-4 [Accessed 22 January 2026].
19. Farchy, J. (2025) 'Uzbekistan top supermarket Korzinka raises $110mn to fund rapid expansion', bne IntelliNews, 12 June. Available from https://www.intellinews.com/uzbekistan-top-supermarket-korzinka-raises-110mn-to-fund-rapid-expansion-385227/ [Accessed 8 April 2026]