Intelligent Blood Supply Management

Demand Forecasting, Donor Segmentation & Recommender System for Inventory Optimization

Key Results: MAPE: **12.14%** | Silhouette Score: **0.41** | Recommender Improvement: **+47%**

1. Problem Background & Motivation

1.1 The Blood Supply Challenge

Blood supply management is one of the most critical challenges in healthcare logistics. Blood banks face a dual challenge: ensuring adequate supply to meet unpredictable demand while minimizing wastage due to the perishable nature of blood products. Unlike conventional inventory management, blood supply must account for unique biological constraints including limited shelf life, compatibility requirements, and the inability to manufacture blood artificially.

Blood shortages lead to delayed surgeries, postponed treatments, and preventable deaths. Conversely, overstocking results in significant wastage due to expiration. According to NBCUS 2019, the United States collects approximately 13 million units of whole blood annually, with wastage rates ranging from 5-15% depending on blood component and facility.

1.2 Critical Challenges

- **Shortage Risk:** Unpredictable demand patterns, seasonal variations, and emergencies create sudden shortages based on surgical schedules, accident rates, and disease outbreaks.
- **Wastage Risk:** Platelets last only 5 days, packed RBCs up to 42 days. Miscalculations lead to expiration and wastage.
- **Blood Type Matching:** 8 major blood types require careful supply management while accounting for compatibility rules.
- **Donor Availability:** Blood banks depend on voluntary donors whose availability varies based on health, schedules, and motivation.

1.3 Research Motivation

A study at B.P. Koirala Institute of Health Sciences (BPKIHS) in Nepal found 7.13% wastage due to expiry. This project applies machine learning to: (1) provide accurate demand forecasts, (2) segment donors based on behavior, (3) create intelligent donor-campaign matching, and (4) detect anomalies in demand patterns.

1.4 Project Objectives

Objective	Technique	Target Metric
Demand Forecasting	SARIMA, Prophet, Random Forest, Gradient Boosting	MAPE < 20%
Donor Segmentation	RFM Analysis + K-Means Clustering	Silhouette Score > 0.35

Recommender System	Hybrid Donor-Campaign Matching	Outperform random baseline
Anomaly Detection	Rolling Z-score Change Point Detection	Detect demand spikes

2. Dataset Description

2.1 Data Sources Overview

This project uses real-world datasets combined with synthetic data based on authoritative sources including NBCUS 2019, WHO blood type distributions, and peer-reviewed research.

Dataset	Records	Type	Source
UCI Blood Transfusion	748	Real	UCI ML Repository
Kaggle Blood Demand	-	Real	Kaggle
Kaggle Blood Donor	-	Real	Kaggle
Synthetic Donor Registry	10,000	Synthetic	NBCUS 2019 Statistics
Demand Time-Series	4,380	Synthetic	3 years daily data
Supply Inventory	4,380	Synthetic	Nepal BPKIHS Study

2.2 UCI Blood Transfusion Dataset

Contains 748 records from a Taiwan mobile blood donation service. This real-world dataset validates our RFM-based segmentation approach.

Column	Description	Statistics
Recency	Months since last donation	Mean: 9.5, Range: 0-74
Frequency	Total number of donations	Mean: 5.5, Range: 1-50
Monetary	Total blood donated (cc)	Mean: 1,379cc, Range: 250-12,500
Time	Months since first donation	Mean: 34.3, Range: 2-98
Donated	Target: donated in March 2007	24% positive class

2.3 Synthetic Donor Registry

10,000 synthetic donor records based on authoritative statistics:

- **Blood Type (WHO):** O+: 37.4%, A+: 28.0%, B+: 20.0%, O-: 6.6%, others: 8.0%
- **Demographics (NBCUS 2019):** 30% first-time donors, 54% male, 15% deferral rate

2.4 Demand Time-Series Dataset

Three years (2021-2023) of daily data (4,380 records) for Whole Blood, Packed RBC, Platelets, and FFP with realistic patterns: 30% lower weekend demand, +15% winter peak, 20-25% holiday spikes.

2.5 Supply Inventory Dataset

Daily inventory tracking with 91.2% average utilization and 8.8% wastage rate.

3. Methodology & Model Choice

3.1 Demand Forecasting Approach

Four forecasting approaches were implemented:

- **SARIMA(1,1,1)(1,1,1,7):** Baseline statistical model for trend and weekly seasonality, parameters chosen via ACF/PACF analysis.
- **Facebook Prophet:** Handles multiple seasonalities and holiday effects with decomposable structure.
- **Random Forest (100 trees):** Captures non-linear relationships and feature interactions.
- **Gradient Boosting:** Sequential tree building to correct previous errors.

3.2 Feature Engineering

Feature	Description	Importance
lag_1	Previous day's demand	0.42 (highest)
rolling_mean_7	7-day rolling average	0.18
lag_7	Same day last week	0.12
rolling_mean_14	14-day rolling average	0.08
day_of_week	Encoded day (0-6)	0.07
rolling_mean_30	30-day rolling average	0.05
month	Month of year (1-12)	0.04
is_weekend	Binary weekend flag	0.04

3.3 Donor Segmentation (RFM + K-Means)

RFM framework combined with K-Means clustering:

- **Recency:** Months since last donation (lower is better)
- **Frequency:** Total number of donations (higher is better)
- **Monetary:** Total volume donated in cc (higher is better)

After StandardScaler normalization, K-Means (k=5) was applied. Optimal k determined via elbow method and silhouette analysis.

3.4 Hybrid Recommender System

Weighted scoring for donor-campaign matching:

```
Score = 0.4 × RFM_Score + 0.2 × Availability + 0.2 × Segment_Match + 0.1 × Blood_Match + 0.1 × Urgency
```

- **RFM Score (40%):** Past behavior predicts future donations
- **Availability (20%):** Eligibility check (not recently donated, no deferrals)
- **Segment Match (20%):** Campaign-segment alignment
- **Blood Type Match (10%):** Priority for needed types
- **Urgency Factor (10%):** Emergency situation boost

3.5 Anomaly Detection

Rolling Z-score with 30-day window and 2.5σ threshold identifies demand values deviating significantly from recent patterns.

4. Experiments & Results

4.1 Demand Forecasting Results

Models trained on 2021-2022, evaluated on 2023 (365 days):

Model	MAE	MAPE	R^2	Training Time
Random Forest	**19.68**	**12.14%**	**0.45**	2.3s
SARIMA	22.19	12.40%	0.20	45.2s
Gradient Boosting	22.11	13.75%	0.28	4.1s
Prophet	24.55	14.82%	0.15	8.7s

Key Findings: Random Forest achieved best performance (MAPE 12.14%, below 20% target). SARIMA performed well despite simplicity (12.40% MAPE). lag_1 contributed 42% of predictive power, indicating strong autocorrelation.

4.2 Feature Importance Analysis

Rank	Feature	Importance	Interpretation
1	lag_1	0.42	Yesterday's demand most predictive
2	rolling_mean_7	0.18	Weekly average captures trends
3	lag_7	0.12	Same weekday pattern matters
4	rolling_mean_14	0.08	Two-week trends provide context
5	day_of_week	0.07	Strong weekly seasonality

4.3 Donor Segmentation Results

K-Means with k=5 achieved Silhouette Score of 0.41 (target: >0.35):

Segment	% Donors	Avg Recency	Avg Frequency	Avg Volume
Champions	11.5%	1.9 mo	12.3	6,150 cc
Loyal	22.3%	3.8 mo	7.2	3,600 cc
Potential	28.5%	4.2 mo	3.5	1,750 cc
At Risk	19.6%	8.4 mo	5.1	2,550 cc
Hibernating	18.1%	12.1 mo	2.8	1,400 cc

Segment Profiles:

- **Champions (11.5%):** Best donors - VIP treatment and retention programs
- **Loyal (22.3%):** Consistent donors - Referral programs
- **Potential (28.5%):** Moderate engagement - Nurturing and conversion
- **At Risk (19.6%):** Disengaging - Urgent reactivation
- **Hibernating (18.1%):** Long-inactive - Last-chance campaigns

4.4 Recommender System Simulation

Comparison across 1,000 simulated campaign iterations:

Strategy	Avg Donations/Campaign	vs Random	Efficiency
Random Selection	~45	baseline	1.0x
RFM-based Only	~58	+29%	1.3x
Hybrid Recommender	**~66**	**+47%**	**1.5x**

Campaign Types:

1. **Winter Blood Drive:** High-frequency donors during December-February peak
2. **O-Negative Emergency:** Universal donor type during shortages
3. **Platelet Donation Week:** Eligible platelet donors (5-day shelf life)
4. **Summer Stock Building:** Proactive collection during low-demand period
5. **Lapsed Donor Reactivation:** At Risk and Hibernating segments

4.5 Anomaly Detection Results

Rolling Z-score identified 47 anomalous days (4.3% of total):

- 15 anomalies: Major holidays (Diwali, Christmas, New Year)
- 12 anomalies: Severe weather events
- 8 anomalies: Major surgical events at partner hospitals
- 12 anomalies: Random/unreported events

5. Insights & Business Interpretation

5.1 Demand Pattern Insights

Pattern	Finding	Business Implication
Weekly Cycle	30% lower weekend demand	Reduce weekend drives, focus weekdays
Winter Peak	+15% December-February	Pre-stock in November
Holiday Spikes	+20-25% major holidays	Maintain emergency reserves
Monday Surge	Highest demand day	Ensure maximum supply Monday AM

5.2 Inventory Management Recommendations

Platelets (12.3% wastage): Just-in-time collection (5-day shelf life), 3-day forecast horizon, rapid donor notification.

Packed RBC (5.2% wastage): Buffer stock strategy (42-day shelf life), 7-day safety stock, FIFO distribution.

O-Negative (6.6% population): Priority collection in all campaigns, separate emergency reserve (48-hour minimum).

5.3 Donor Engagement Strategies

Segment	Strategy	Expected Outcome
Champions	VIP programs, early access, recognition	95% retention
Loyal	Referral incentives, milestone celebrations	+15% referrals
Potential	Education, convenience improvements	40% convert to Loyal
At Risk	Personalized outreach, feedback surveys	25% reactivation
Hibernating	Final reactivation attempt, then archive	10% recovery

5.4 Projected Business Impact

Wastage Reduction:

- Current wastage: 8.8% (BPKIHS benchmark)
- Projected with forecasting: 5.0%
- Estimated annual savings: ~$150,000 (medium-sized blood bank)

Collection Efficiency:

- 47% improvement in donation yield from recommender
- Reduced cost per unit through targeted outreach
- Better matching of collected types to demand

Operational Benefits:

- Automated forecasting reduces manual planning
- Data-driven donor management improves efficiency
- Anomaly detection provides early warning for surges

6. Limitations & Future Scope

6.1 Current Limitations

Data Limitations:

- Synthetic data may not capture all real-world complexities
- Real-world validation needed before deployment
- Geographic and cultural variations not fully modeled

Model Limitations:

- Emergency situations and mass casualty events unpredictable
- External factors (epidemics, natural disasters) not incorporated
- Donor behavior models assume rational response to outreach

Operational Limitations:

- Integration with existing blood bank systems required
- Staff training and change management not addressed
- Regulatory compliance varies by jurisdiction

6.2 Future Research Directions

Advanced Deep Learning:

- LSTM networks for long-term temporal dependencies
- Transformer models for multi-variate forecasting
- Graph Neural Networks for donor-hospital relationships

Reinforcement Learning:

- Dynamic inventory optimization using RL agents
- Adaptive campaign timing based on real-time feedback
- Multi-objective optimization (supply, cost, wastage)

IoT and Real-time Monitoring:

- Temperature and condition monitoring of stored products
- Real-time inventory tracking across facilities
- Predictive maintenance for storage equipment

Multi-center Optimization:

- Network-level optimization across branches
- Transfer optimization between facilities
- Regional coordination during shortages

7. Conclusion

This project demonstrates the application of advanced machine learning to blood supply management. Through demand forecasting, donor segmentation, and intelligent recommendation, we show ML can significantly improve blood bank operations.

7.1 Achievement of Objectives

Objective	Target	Achieved	Status
Demand Forecasting	MAPE < 20%	12.14%	**Exceeded**
Donor Segmentation	Silhouette > 0.35	0.41	**Exceeded**
Recommender System	Beat random	+47%	**Exceeded**
Anomaly Detection	Detect spikes	4.3% identified	**Achieved**

7.2 Key Contributions

1. **Integrated ML Pipeline:** End-to-end solution combining multiple techniques
2. **Feature Engineering:** lag_1 contributes 42% of predictive power
3. **Actionable Segmentation:** Interpretable segments with specific strategies
4. **Hybrid Recommender:** Outperforms RFM-based approach by 18%
5. **Business Insights:** Technical findings translated to operational recommendations

7.3 Practical Implications

- ML forecasting can reduce wastage from 8.8% to 5%, saving resources
- Targeted outreach increases donation yield by up to 47%

• Anomaly detection enables proactive response to demand surges
• Combined techniques create synergies amplifying individual benefits

> *"Advanced ML is not about prediction accuracy alone. It is about discovering structure, extracting insights, and supporting decision-making in complex data."*

7.4 Final Remarks

Blood supply management is a domain where ML can have life-saving impact. By improving collection, storage, and distribution efficiency, these techniques help ensure blood availability while minimizing waste. This project provides a foundation for blood banks to adopt data-driven operations.

References:

1. NBCUS 2019, US Dept. of Health and Human Services
2. Singh et al. (2023). Blood wastage at BPKIHS. J Pathology Nepal
3. UCI ML Repository - Blood Transfusion Dataset
4. WHO - Blood Safety and Availability
5. Kaggle Blood Demand and Donor Datasets

Code: 5 Jupyter notebooks | Visualizations: data/visualizations/ | Data: Google Drive