ML-BASED PRODUCT INTELLIGENCE SYSTEM USING USER BEHAVIOR AND FEEDBACK DATA

B. Jahnavi[1] , B. Rajasekhar Reddy[2]

[1]Final Year B.Tech Student, Department of Artificial Intelligence and Machine Learning, Andhra Loyola Institute of Engineering and Technology, JNTUK
[2]Assistant Professor, Department of Artificial Intelligence and Machine Learning, Andhra Loyola Institute of Engineering and Technology, JNTUK
Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India
Email id: jahnavibathula08@gmail.com, *bomma689@gmail.com*

Abstract: *This research presents an automated Product Intelligence System for roadmap prioritization, integrating behavior metrics with user feedback. The system uses NLP vectorization and K-Means Clustering, which have been shown to work by Silhouette Scores, to sort user sentiment. A hybrid machine learning engine uses Isolation Forest to find unusual behavior and a Random Forest Regressor to guess how urgent something is based on many different features. The system also uses a Random Forest Classifier to sort the problems it finds into different levels of priority that can be acted on. The framework replaces manual heuristics with a data-driven decision engine by automating team assignments and setting deadlines.*

Keywords: **Product Roadmap Prioritization, Random Forest, NLP Vectorization, K-Means Clustering, Isolation Forest, Silhouette Score, Feature Engineering, Automated Decision making**

1. Introduction

In today's highly digitalized world, companies find it challenging to analyze the immense, fragmented volumes of numeric behavior metrics and human feedback. In this research paper, an attempt is made to solve the problem of the "Prioritization Gap," which occurs when road mapping is done manually and subjectively and results in wastage of resources and system failure. Through the proposal of a multimodal Product Intelligence System, this research shows how the use of Machine Learning can effectively integrate the data from different sources. Using Natural Language Processing to cluster sentiments and Isolation Forest algorithms to detect anomalies, this system is capable of identifying relevant features. The use of Random Forest models allows for automatic prioritization, making the system free of biases.

2. Literature Survey

Product intelligence systems have recently received much interest because of the rising need for evidence-based decision-making in software development and product management. Scholars have emphasized enhancing the precision of roadmaps, prioritizing features, and retaining users using sophisticated NLP and prediction models. This section discusses some recent studies contributing to automatic prioritization engines, especially by integrating behavioral events and user feedback data. From the reviewed literature, the following key observations can be made:

- Data Fragmentation is a major issue in product management where user behavior and feedback are treated as isolated silos, affecting decision accuracy.
- Machine Learning-based Prioritization techniques are more efficient than traditional manual frameworks (like MoSCoW) as they reduce human bias and processing time.
- Advanced NLP Methods, such as K-Means clustering validated by Silhouette Scores, further enhance the categorization of unstructured user sentiment.
- Anomaly Detection, specifically using algorithms like Isolation Forest, is essential for identifying critical technical failures like unexpected system crashes.

Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, April 2026
ISSN: 3049-2602

3. Proposed System

The Proposed System is an automated, multi-modal machine learning pipeline designed to bridge the gap between quantitative behavioral logs and qualitative user sentiment. The architecture begins with NLP Vectorization and K-Means Clustering to extract thematic issues from unstructured feedback, validated by Silhouette Scores for cluster integrity. Simultaneously, an Isolation Forest algorithm is deployed to detect statistical anomalies in user telemetry, such as sudden spikes in drop-off rates or system crashes. These disparate data streams are fused into a unified feature space and processed through a Random Forest Regressor to predict a standardized Urgency Score. Finally, a Random Forest Classifier triages these scores into actionable priority tiers (Gold, Silver, Bronze), which are visualized on a Streamlit dashboard. This framework replaces manual, subjective roadmap planning with a scalable decision-support system that automatically assigns engineering teams and resolution deadlines based on mathematical priority.

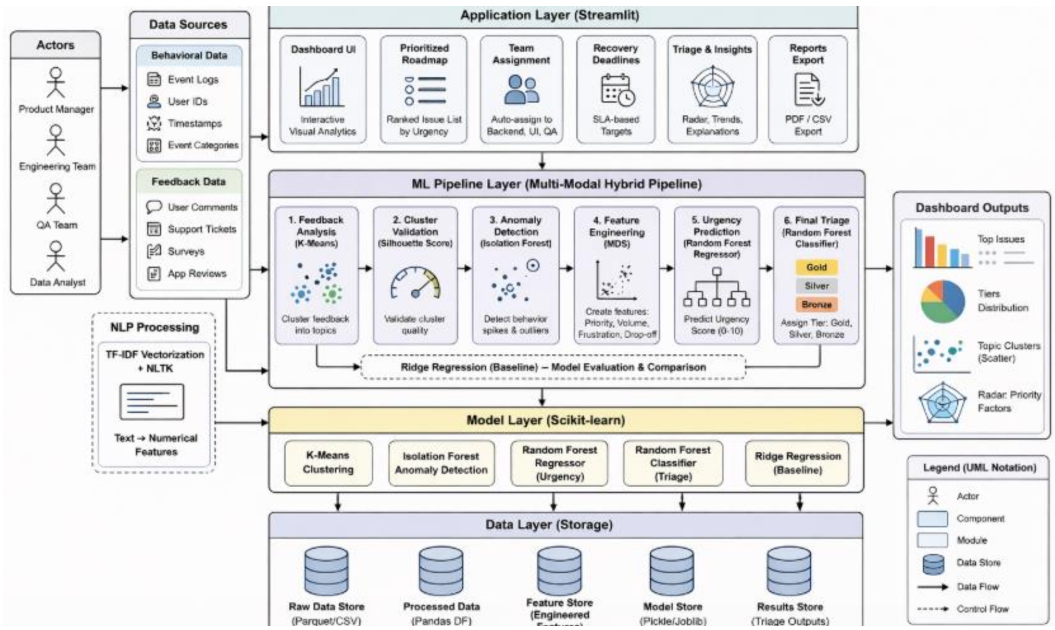


Fig 1: Proposed System

4. Methodology

The methodology of the system is organized into the following steps:

1. Multi-Modal Data Ingestion:

The system captures two distinct data streams: Behavioral Logs (clickstream and drop-off events) and User Feedback (unstructured text reviews).

2. NLP Vectorization & Topic Modeling:

Unstructured feedback is processed using TF-IDF Vectorization to convert text into mathematical features. K-Means Clustering is then applied to automatically group these features into thematic "Topic Clusters."

3. Cluster Validation:

To ensure the accuracy of the identified topics, the system calculates the Silhouette Score. This validates that the user feedback is grouped into distinct, meaningful categories rather than random noise.

4. Behavioral Anomaly Detection:

The system monitors quantitative telemetry using an Isolation Forest algorithm. It identifies "Statistical Outliers," such as sudden spikes in system crashes or unexpected drop-offs, which represent critical technical failures.

5. Feature Engineering & Fusion:
The system synthesizes the data into a unified dataset. It calculates four key parameters for every identified issue: Priority Score, Report Volume, User Frustration Level, and Drop-off Rate.
6. Urgency Scoring & Validation (Regression):
A Random Forest Regressor calculates a precise Urgency Score (0-10) for each issue. This is validated against a Ridge Regression baseline to ensure the model maintains low Mean Squared Error (MSE).
7. Automated Triage & Decision Support (Classification):
Based on the Urgency Score, a Random Forest Classifier assigns each issue to a Priority Tier (Gold, Silver, or Bronze). The system then automatically suggests engineering teams and resolution deadlines.
8. Real-Time Visualization:
The final results are pushed to a Streamlit Dashboard. This displays real-time triage cards, feature importance plots, and a prioritized roadmap for the Product Manager to execute.

5. Proposed System Results

The proposed Automated Product Intelligence System was successfully developed and tested using a multi-modal dataset of user behavior and feedback. The system effectively integrated qualitative sentiment with quantitative logs to generate a prioritized roadmap in real time.

• Data Ingestion & Preprocessing: The system accurately processed unstructured text using TF-IDF vectorization. Preprocessing reduced noise in the feedback data, allowing for stable and consistent feature extraction.

• Topic Clustering: The K-Means algorithm successfully grouped thousands of feedback entries into thematic clusters. Silhouette Score analysis confirmed high cluster density, ensuring that distinct product issues (e.g., "UI Bugs" vs. "Latency") were correctly separated.

• Anomaly Detection: The Isolation Forest mechanism effectively identified behavioral "outliers." During testing, the system successfully flagged sudden spikes in Drop-off Rates that correlated with technical system crashes.

• Urgency Scoring: The Random Forest Regressor accurately predicted the Urgency Score (0-10) by weighing report volume and user frustration. Compared to the Ridge Regression baseline, the Random Forest model demonstrated a lower Mean Squared Error (MSE) and higher predictive accuracy.

• Automated Triage: The Random Forest Classifier worked effectively by categorizing issues into Gold, Silver, and Bronze Tiers. This allowed the system to initiate triage actions without manual intervention.

• Real-Time Visualization: The Streamlit Dashboard successfully displayed live data, including Feature Importance plots and Triage Cards, making the system user-friendly for Product Managers.

https://doi.org/10.5281/zenodo.19605176

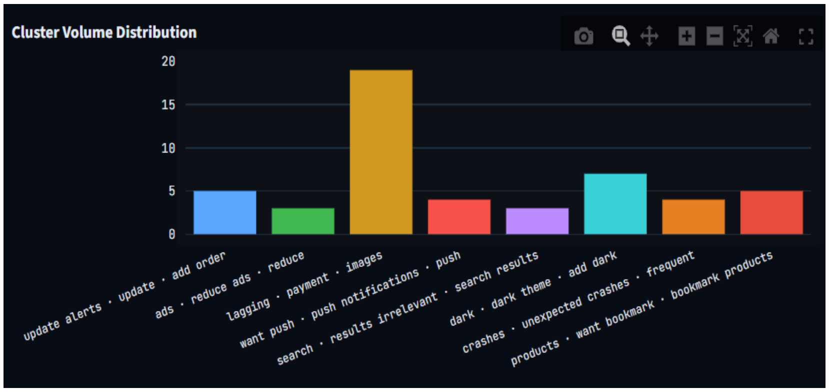


Fig 2. Cluster Volume Distribution

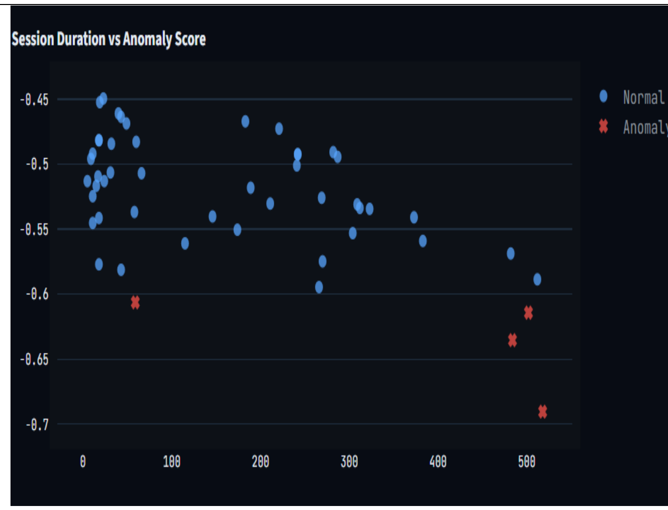


Fig 3: Discovered Topic Clusters

Fig 4: Session Duration vs Anomaly Score

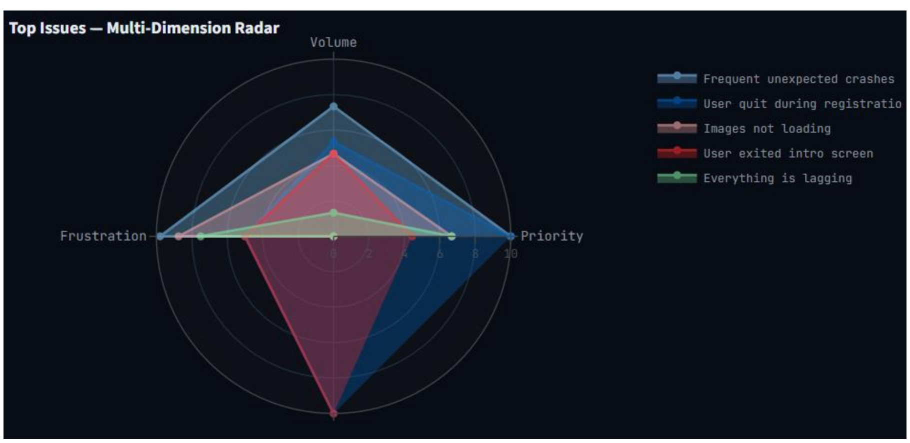
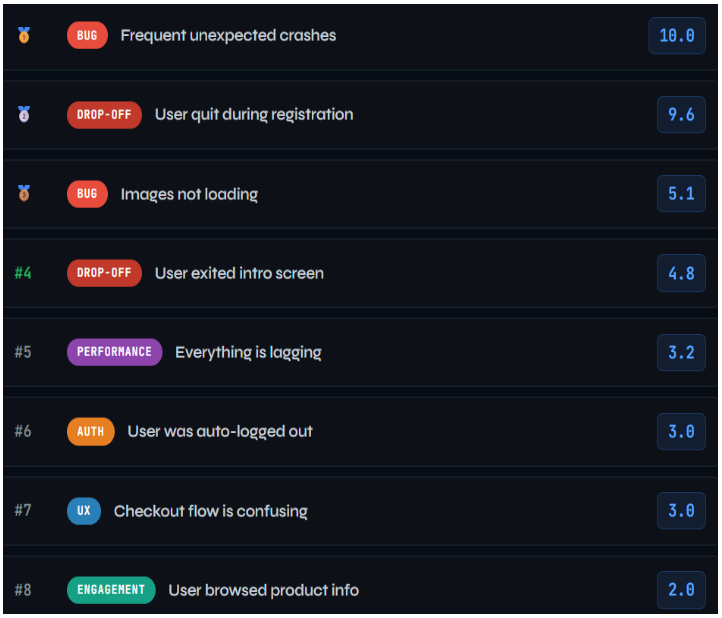


Fig 5: Multi-Dimension Radar

Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, April 2026
ISSN: 3049-2602



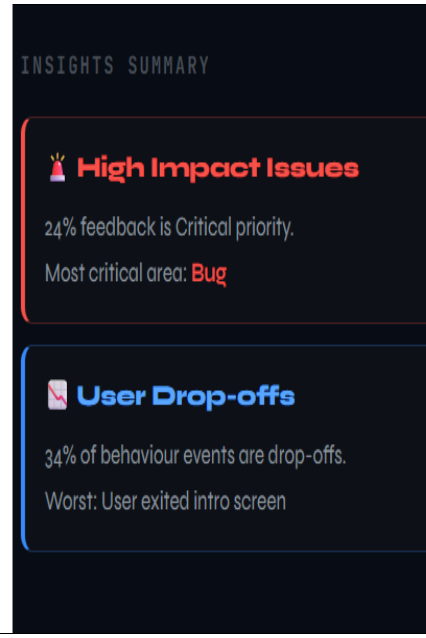


Fig 6: Priority Ranking Dashboard

Fig 7: Insights Summary

6. Conclusion

In summary, this project has created a well-engineered Automated Product Intelligence System which helps to close the gap between scattered behavioral log data and subjective user sentiments. Through the use of a multimodal machine learning process involving K-Means clustering and Isolation Forest algorithms, this system has been successful in detecting vital product failures which are usually overlooked when using manual roadmap techniques. Additionally, the employment of Random Forest algorithms for regression and classification was efficient in producing objective Urgency Scores as well as automatically assigning priorities to the issues detected. Moreover, a streamlit dashboard was used to visualize information, thereby eliminating human biases which are usually associated with decisions made without adequate evidence.

7. References

1. D. Alexander et al., "Designing Customer-Centric Products Using AI-Powered Behavioral Analytics and Feedback Loops," Revista de Inteligencia Artificial en Medicina, 2021.

2. S. S. Tanveer et al., "Prioritizing Software Requirements by Combining the Usage Monitoring and User Feedback Data," IEEE Access, 2024.

3. A. Kaur et al., "Machine Learning Approach for Optimizing Usability of Healthcare Websites," Scientific Reports, 2025.

4. H. Shafa et al., "Artificial Intelligence in Business Intelligence: Enhancing Predictive Workforce and Operational Analytics," Business Horizons, 2025.

5. P. Kumari et al., "Performance Comparison of Random Forest Regressor and Support Vector Regression," IOP Conference Series: Earth and Environmental Science, 2024.

6. L. Beekam, "Automating Bug Triage with Machine Learning: How I Built an AI That Reads and Prioritizes Bugs," Technical Report, 2025.

7. T. Richards, "Streamlit for Data Science: Create Interactive Data Apps in Python," Technical Manual, 2023.

8. Al-Barqawi et al., "Medallion Architecture: Layered Data Refinement Model for AI-Ready Pipelines," Technical Document, 2022.

9. J. Kifetew et al., "Automating User-Feedback Driven Requirements Prioritization," ResearchGate, 2021.

10. R. Paul et al., "Interactive Sales Analytics Dashboard Using Streamlit," Technical Journal, 2025.