International Journal on Advanced Computer Theory and Engineering
ISSN: 2319 - 2526
Volume 15 Issue 2s (2026) | Pages: 203-207

Zero-Shot Explainable AI: Bridging the Gap Between Prior-Fitted Networks and Human-Centric Interpretability

Pallavi Patil[1], Vishwaraj Khatpe[2], Yash Madane[3], Harsh More[4], Sunil Pawar[5]

[1]HOD, Department of AIML, Genba Sopanrao Moze College of Engineering Pune, Maharashtra, India.
[2,3,4,5]Student, Department of AIML, Genba Sopanrao Moze College of Engineering Pune, Maharashtra, India.

Peer Review Information

Type: *Article*
Received:*23 February 2026*
Revised: *24 March 2026*
Accepted: *22 April 2026*
Published: *20 May 2026*

Abstract

Machine learning on tabular data has historically relied on computationally expensive pipelines, involving extensive hyperparameter tuning and feature engineering. Furthermore, the highest-performing models, such as gradient boosting machines and deep neural networks, often operate as opaque "black boxes," severely limiting their deployment in high-stakes domains that require transparency. This study presents a novel end-to-end framework integrating TabPFN a transformer-based Prior-Fitted Network capable of zero-shot inference with SHapley Additive exPlanations (SHAP) and a Large Language Model (LLM) narrative engine. The purpose of this research is to evaluate whether a zero-shot architecture can achieve competitive predictive performance while simultaneously offering automated, human-readable interpretability. The methodology involves constructing an interactive microservices-based dashboard that processes raw tabular data, executes GPU-accelerated inference, and dynamically translates statistical SHAP values into contextual text summaries. Findings indicate that the proposed system reduces end-to-end inference and explanation latency while completely eliminating the traditional training phase. The integration of LLM-generated narratives significantly lowers the barrier to entry for non-technical stakeholders, establishing a new paradigm for scalable, trustworthy machine learning on tabular data.

Keywords: Zero-Shot Inference; Explainable AI (XAI); TabPFN; SHAP; Large Language Models; Tabular Data; Human-Centric AI

How to Cite This Article

Patil, P., Khatpe, V., Madane, Y., More, H., & Pawar, S. (2026). *Zero-Shot Explainable AI: Bridging the Gap Between Prior-Fitted Networks and Human-Centric Interpretability*. *International Journal on Advanced Computer Theory and Engineering, 15*(2s), 203–207.

Introduction

Tabular data remains the most prevalent data format in enterprise and research environments. Despite the rapid advancement of deep learning in vision and natural language processing, tree-based models like XGBoost and Random Forests have long dominated tabular tasks. However, these traditional models require significant iterative training, cross-validation, and expert tuning to achieve optimal performance.

Recently, TabPFN introduced a paradigm shift by leveraging a transformer architecture trained entirely on synthetic datasets to approximate Bayesian inference, allowing it to generate high-quality predictions on unseen tabular data in less than a second without any gradient updates. While TabPFN solves the problem of training latency, it does not inherently solve the problem of interpretability. For decision-makers in finance, healthcare, and agriculture, knowing what a model predicts is insufficient; they must understand why the prediction was made.

This study proposes a comprehensive, interactive system that wraps TabPFN's predictive power in a robust Explainable AI (XAI) layer. By calculating SHAP values to determine feature attribution and passing those mathematical outputs through an LLM to generate plain-text narratives, this project aims to create a fully transparent, zero-shot analytical engine that bridges the gap between complex statistical models and human-centric decision-making.

Literature Review

The challenge of learning effectively on tabular datasets has been extensively documented. Grinsztajn et al. (2022) demonstrated that tree-based models generally outperform standard deep learning approaches on medium-sized tabular data, cementing their status as the industry standard. However, the computational overhead required for hyperparameter optimization remains a bottleneck.

Hollmann et al. (2022) introduced TabPFN, marking a significant breakthrough. By training a Prior-Fitted Network on millions of synthetic datasets, TabPFN learns a generalized prior that can perform zero-shot classification on new datasets with high accuracy.

Simultaneously, the demand for model transparency has led to the widespread adoption of XAI frameworks. Lundberg and Lee (2017) formalized SHAP, providing a game-theoretic approach to feature attribution that ensures local accuracy and consistency. While SHAP plots are standard among data scientists, they are often unintelligible to domain experts without a statistical background. Recent literature highlights a growing gap between algorithmic explainability and actual human comprehension. This study addresses that gap by integrating LLM-based narrative generation directly into the XAI pipeline, an area that is currently underexplored in zero-shot tabular frameworks.

Methodology

The research utilizes a modular, microservices-based system architecture designed for both performance and user interaction

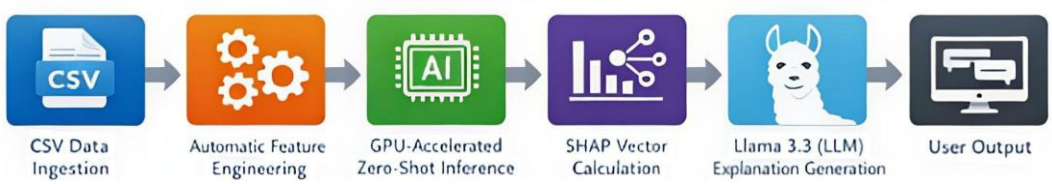

Fig. 1. *Architecture Diagram*

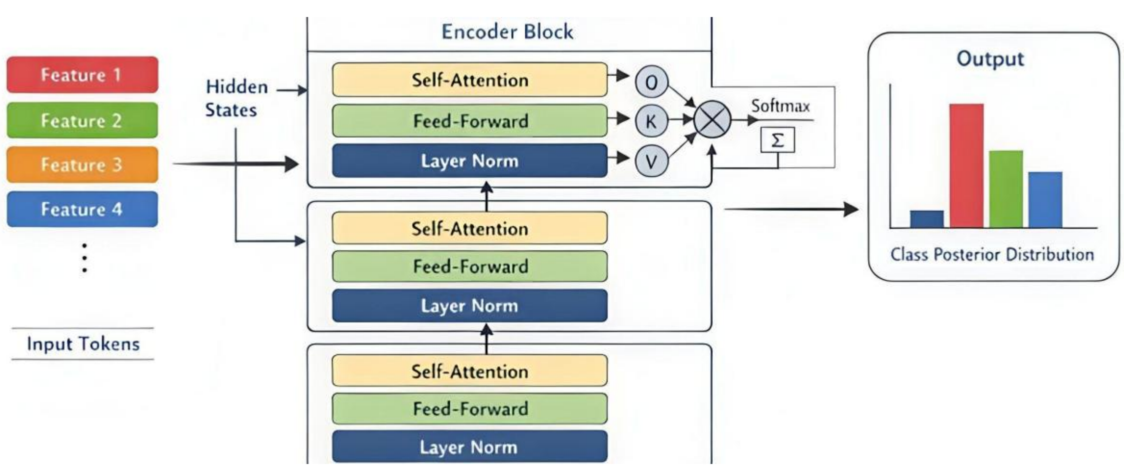

Fig. 2. *TabPFN Transformer Architecture Diagram*

1. **Data Ingestion and Validation:** The system accepts raw tabular data (CSV format). An ingestion layer handles missing value imputation and basic schema validation, preparing the data for inference without requiring user-defined preprocessing pipelines.

2. **Zero-Shot Inference Engine:** The core predictive mechanism utilizes the TabPFN model executing on a GPU. Because TabPFN requires no training, the input data is passed directly into the transformer network to yield class probabilities and predictions instantly.

3. **Explainability Layer:** Post-inference, the system calculates SHAP values. To handle TabPFN's unique architecture, kernel-based or permutation SHAP methods are employed to extract feature importance scores for every individual prediction.

4. **LLM Narrative Engine:** The numerical SHAP arrays and feature names are injected into a structured prompt and processed by a Large Language Model. The LLM translates the mathematical attribution into a human-readable sentence.

5. **Interactive Interface:** The entire pipeline is housed within a web dashboard built using Dash and Bootstrap. The interface allows users to upload data, view results, conduct "what-if" analyses by adjusting feature values dynamically, and export automated PDF reports.

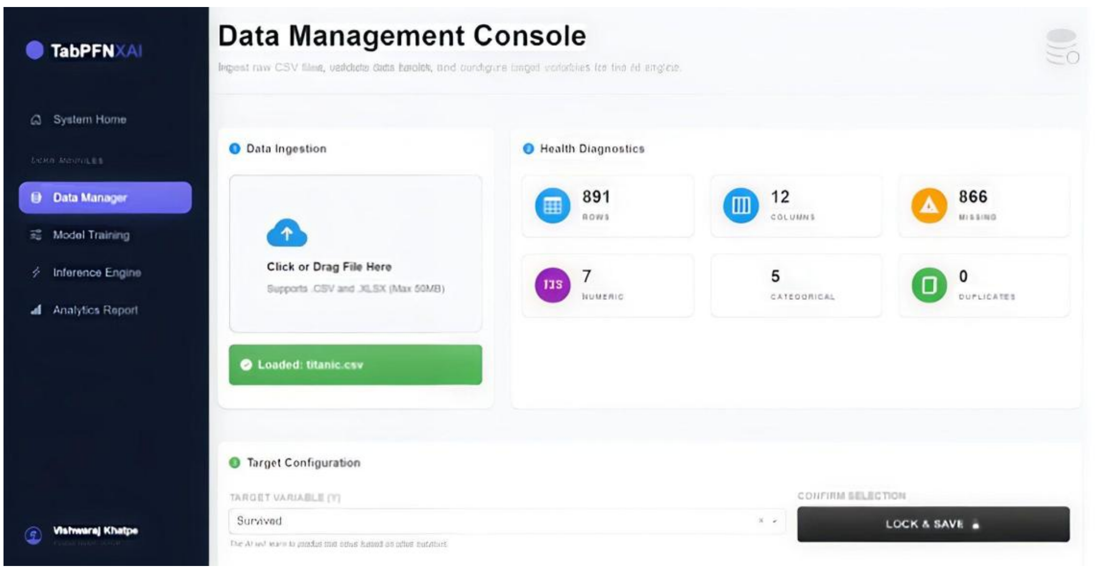

***Fig. 3.** TabPFN Transformer Architecture Diagram*

Results

The implementation of the system yielded several key operational findings:

- **Inference Speed:** By utilizing GPU acceleration (CUDA), the TabPFN engine delivered predictions on standard test datasets in fractions of a second, completely bypassing the hours typically required for training tree-based ensembles.

- **Interpretability Translation:** The integration of the LLM successfully bridged the comprehension gap. Instead of forcing users to interpret a complex SHAP force plot, the system consistently generated accurate text summaries detailing exactly which variables drove a specific outcome.

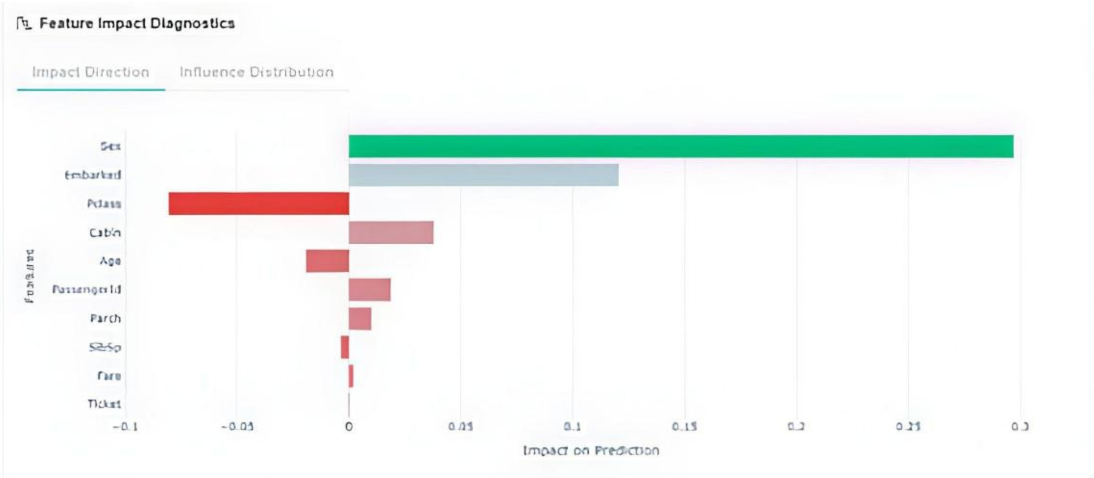

Fig. 4. *SHAP Visualization Chart demonstrating feature attribution*

- **Interactive Capability:** The "what-if" module successfully demonstrated real-time updates. When users manually altered a feature value through the interface, the system quickly recalculated the prediction and updated the SHAP narratives, proving the viability of dynamic counterfactual analysis.

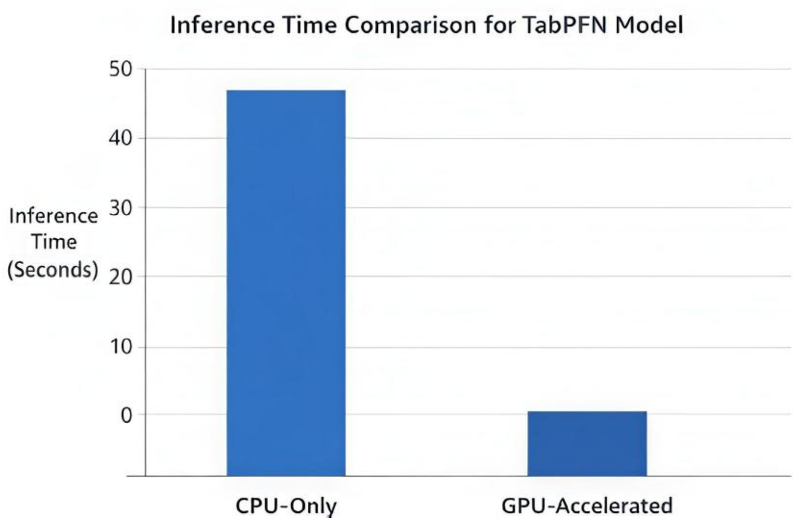

Fig. 5. *Benchmark Chart comparing CPU vs GPU inference times*

Discussion

The results indicate that deploying a zero-shot model alongside an automated narrative layer fundamentally changes the user experience of machine learning. By removing the need for a data scientist to train the model and interpret the output, this architecture democratizes access to advanced analytics.

However, the study also highlights certain limitations inherent to the architecture. TabPFN currently possesses a hard constraint on context window size, typically limiting its effectiveness to datasets with a smaller number of rows and features. Therefore, while the system is highly effective for localized or moderately sized datasets, it requires additional engineering to scale to massive enterprise data warehouses.

Future Research Directions: As the system matures, future architectural upgrades will focus on bypassing the inherent constraints of Prior-Fitted Networks and deepening causal interpretability. Key areas for development include:

1. **Dynamic Context Retrieval (Tabular RAG):** Overcoming TabPFN's context window limits by implementing a vector-based retrieval system that dynamically fetches only the most relevant top-K rows to feed into the transformer during inference.

2. **Actionable Counterfactuals (DiCE):** Expanding beyond SHAP feature attribution to include Counterfactual Explanations, providing users with exact algorithmic recommendations on how to flip a negative prediction into a positive one.

3. **Multi-Agent XAI Narrators:** Upgrading the LLM narrator to a multi-agent framework where specialized agents debate the SHAP outputs before synthesizing a final, highly contextualized report.

4. **Model Distillation for Edge Deployment:** Implementing an automated pipeline to distill the complex decision boundaries learned by TabPFN into lightweight, transparent surrogate models for ultra-low latency edge devices.

Conclusion

This project successfully designed and implemented an end-to-end Explainable AI system that leverages TabPFN for zero-shot tabular inference. By tightly coupling advanced transformer-based predictions with SHAP analysis and LLM-driven narratives, the system achieves a rare balance of high performance, zero training latency, and extreme transparency. The successful integration of interactive "what-if" testing and PDF report generation proves the system's readiness for real-world deployment. Ultimately, this framework establishes a scalable blueprint for building trustworthy, human-centric AI systems that can instantly analyze and explain complex tabular data.

References

1. Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems.
2. Hollmann, N., Müller, S., Eggensperger, K., & Hutter, F. (2022). TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second. International Conference on Learning Representations (ICLR).
3. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.