Research Proposal

Foundation Models for Tabular & Time-Series Data

Submitted to IITB-Optiver AI Innovation Lab

Principal Investigator Information

PI Name:	Prof. Siddhartha Duttagupta
Department:	Electrical Engineering, IIT Bombay
Email:	sdgupta@ee.iitb.ac.in
Contact No:	+91-XXXXXXXXXX

Project Title:	Foundation Models for Tabular & Time-Series Data
Research Themes:	2.1, 2.2, 2.3, 2.4, 2.5 (Foundation Models)
Project Duration:	6 months
Estimated Budget:	₹13,50,000 (Thirteen Lakhs Fifty Thousand)

Project Summary

This project aims to design and study Transformer-based foundation models for structured and time-series data, with a primary focus on financial and enterprise transaction records. Conventional machine learning approaches for tabular data depend heavily on manual feature engineering and task-specific modeling, which restricts scalability and cross-domain generalization. The proposed research leverages BERT-style Transformer architectures to model sequential and structured data directly, enabling the learning of temporal and contextual patterns through self-supervised training.

The project further explores representation learning techniques for tabular data to create reusable embeddings that can be efficiently adapted to downstream tasks with limited labeled data. Where applicable, multimodal learning is investigated by integrating structured records with textual and visual information using unified Transformer-based representations. In addition, the research studies data-driven segmentation and pattern discovery methods to identify meaningful groups and behaviors without predefined labels. The expected outcome is a scalable and reusable modeling framework that supports reliable risk prediction, anomaly detection, and behavioral analysis across multiple domains, contributing both academically and practically to the advancement of structured data modeling.

Theme 2: Foundation Models for Tabular & Time-Series Data

Problem Statement

While transformer-based foundation models like BERT and GPT have revolutionized natural language processing, financial institutions continue to build task-specific models from scratch for each application. Financial data is primarily structured (tables and transaction logs) rather than unstructured text, yet no general-purpose foundation models exist for this domain. Each bank, insurance company, and fintech firm independently develops models for credit scoring, fraud detection, and risk assessment, resulting in duplicated efforts and inability to leverage cross-domain patterns.

Research Objectives

1. **Transformer Architectures for Transaction Logs (2.1):** Build efficient transformer architectures specifically optimized for high-frequency financial transactions, capable of processing millions of transaction sequences with positional encoding for temporal patterns.

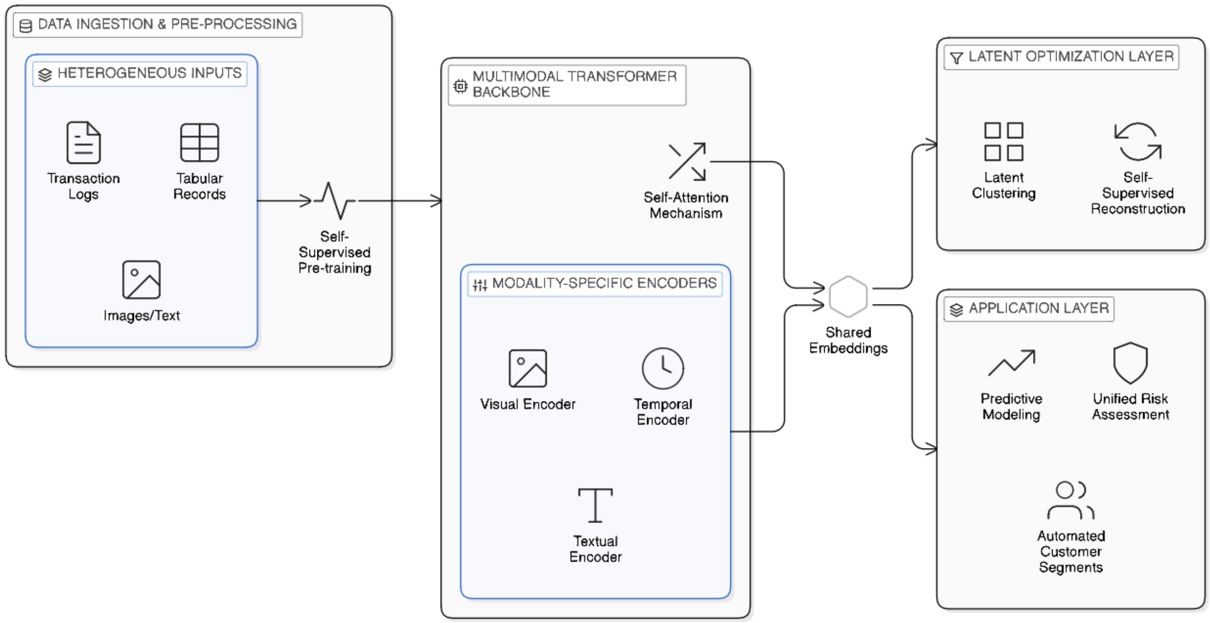
2. **Pretrained Financial Embeddings (2.2):** Create universal embedding layers for structured financial data analogous to BERT embeddings for text, enabling transfer learning across different financial applications and institutions.

3. **Cross-Industry Foundation Models (2.3):** Develop a single adaptable model that generalizes across credit, insurance, retail, and healthcare sectors with minimal fine-tuning, learning industry-agnostic temporal and behavioral patterns.

4. **Multimodal Fusion (2.4):** Enable joint modeling of heterogeneous financial data including contracts (text), claims/transactions (time series), and KYC documents (images) through unified multimodal representations.

5. **SAM-like Segmentation for Tabular Data (2.5):** Implement data-driven customer and transaction segmentation without predefined labels, automatically discovering behaviorally consistent groups through self-supervised learning.

Methodology



The proposed methodology aims to build foundation models for tabular and time-series data using Transformer-based architectures, designed to operate at scale on real-world financial datasets. The work begins with the development of efficient Transformer architectures optimized for high-frequency transaction logs, where financial events are represented as ordered sequences with explicit temporal and positional encodings. These models are designed to capture long-range dependencies, periodic behavior, and anomalous patterns across large volumes of transactional data.

To enable reusable learning across tasks and institutions, the methodology incorporates self-supervised pretraining to create universal embedding representations for structured financial data, analogous to BERT embeddings for text. Masked feature prediction and sequence reconstruction objectives are used to learn generalizable representations from unlabeled tabular and temporal data. These pretrained embeddings serve as a foundation that can be adapted to downstream tasks with minimal labeled data.

The research further proposes the development of cross-industry foundation models by training shared Transformer encoders on heterogeneous datasets spanning credit, insurance, retail, and healthcare domains. This approach enables the model to learn industry-agnostic temporal and behavioral patterns, while lightweight task-specific heads allow efficient fine-tuning for domain-specific applications.

Multimodal fusion is incorporated by jointly modeling structured transaction data with unstructured inputs such as textual contracts and visual KYC documents. Modality-specific encoders are combined using attention-based fusion mechanisms to produce unified representations that preserve cross-modal relationships. Finally, the methodology includes SAM-inspired segmentation techniques applied to learned embeddings, enabling data-driven discovery of customer and transaction segments without predefined labels. Model performance is systematically evaluated for generalization, scalability, and robustness across datasets and tasks.

Expected Outcomes

- Development of a Transformer-based foundation modeling framework for structured and time-series data, capable of learning temporal and contextual patterns directly from raw inputs.

- Creation of pretrained, self-supervised representations for tabular data to reduce dependence on manual feature engineering and labeled datasets.

- A unified modeling approach for integrating structured data with textual and visual information through multimodal Transformer architectures.

- Automated segmentation and pattern discovery mechanisms using learned embeddings to identify meaningful groups and behavioral patterns.

- Empirical validation of the proposed methods on real-world public datasets, with reproducible and well-documented results.

- Research outputs suitable for academic dissemination and further extension in applied industrial settings.

Project Deliverables and Timeline

Phase 1: Months 1–2 (Foundation & Preparation)

- Comprehensive literature review covering Transformer models for tabular, time-series, and multimodal data
- Identification of relevant public financial and transactional datasets
- Design of data collection, cleaning, and preprocessing pipelines
- Establishment of baseline models and evaluation metrics for comparison

Phase 2: Months 3–4 (Model Development)

- Implementation of Transformer-based models for structured and time-series data using BERT-style architectures
- Self-supervised pretraining of tabular representations on selected datasets
- Initial experiments on segmentation and pattern discovery using learned embeddings
- Performance benchmarking against baseline methods

Phase 3: Months 5–6 (Integration & Validation)

- Integration of developed components into a unified modeling framework
- Experimental validation across multiple datasets and tasks
- Performance analysis, ablation studies, and result interpretation
- Documentation of methodology, experimental results, and limitations
- Preparation of final project report and draft research paper for submission

Final Deliverables (End of Month 6)

- A working prototype of a Transformer-based foundation model for structured and time-series data
- Pretrained embeddings and trained model checkpoints
- Reproducible experimental results with clear evaluation metrics
- Well-documented codebase and final technical report

Budget Breakdown

Category	Amount (₹)	Percentage
Research Staff (PhD/Masters students)	5,30,000	41.7%
Computing Infrastructure (GPU/Cloud)	3,80,000	29.2%
Data Acquisition and Licensing	2,30,000	16.7%
Travel and Conference Participation	1,30,000	8.3%
Miscellaneous and Contingency	80,000	4.1%
Total	**13,50,000**	**100%**

Research Team

Principal Investigator:

Prof. Siddhartha Duttagupta, Department of Electrical Engineering, IIT Bombay

Co-Investigators:

Prof. Subhash Kasture, Managing Lead, FAAIR Organization, Aabid Hassan

Research Staff / Students Involved:

Vrushika K Panchal
B.Tech. in Data Science,
Research and Development, FAAIR Organization

Yash K Dhasal
B.Tech. Computer Science & Engineering,
Research and Development, FAAIR Organization

Conclusion

This project presents a systematic approach to developing Transformer-based foundation models for structured and time-series data, with particular emphasis on financial and enterprise transaction records. By adopting BERT-style Transformer architectures and self-supervised learning strategies, the proposed work seeks to overcome key limitations of conventional tabular data modeling methods, including heavy dependence on manual feature engineering, limited scalability, and weak cross-domain generalization. The incorporation of temporal modeling and multimodal data integration further enhances the capability of the framework to handle complex, real-world data scenarios.

In addition, the project explores data-driven segmentation and pattern discovery techniques that enable meaningful interpretation of learned representations without reliance on predefined labels. Through careful experimental validation on public datasets and rigorous performance analysis, the research aims to deliver reproducible and extensible outcomes. Overall, this work is expected to contribute both methodological advancements and practical insights, supporting future academic research and real-world applications in areas such as risk prediction, anomaly detection, and behavioral analysis.

Contact: sdgupta@ee.iitb.ac.in