



PROBLEM STATEMENT –3

PREDICTING COMPLETION OF CLINICAL STUDIES WITH EXPLAINABILITY

Presented by
Yuvika Mishra
Siddhant Nijhawan

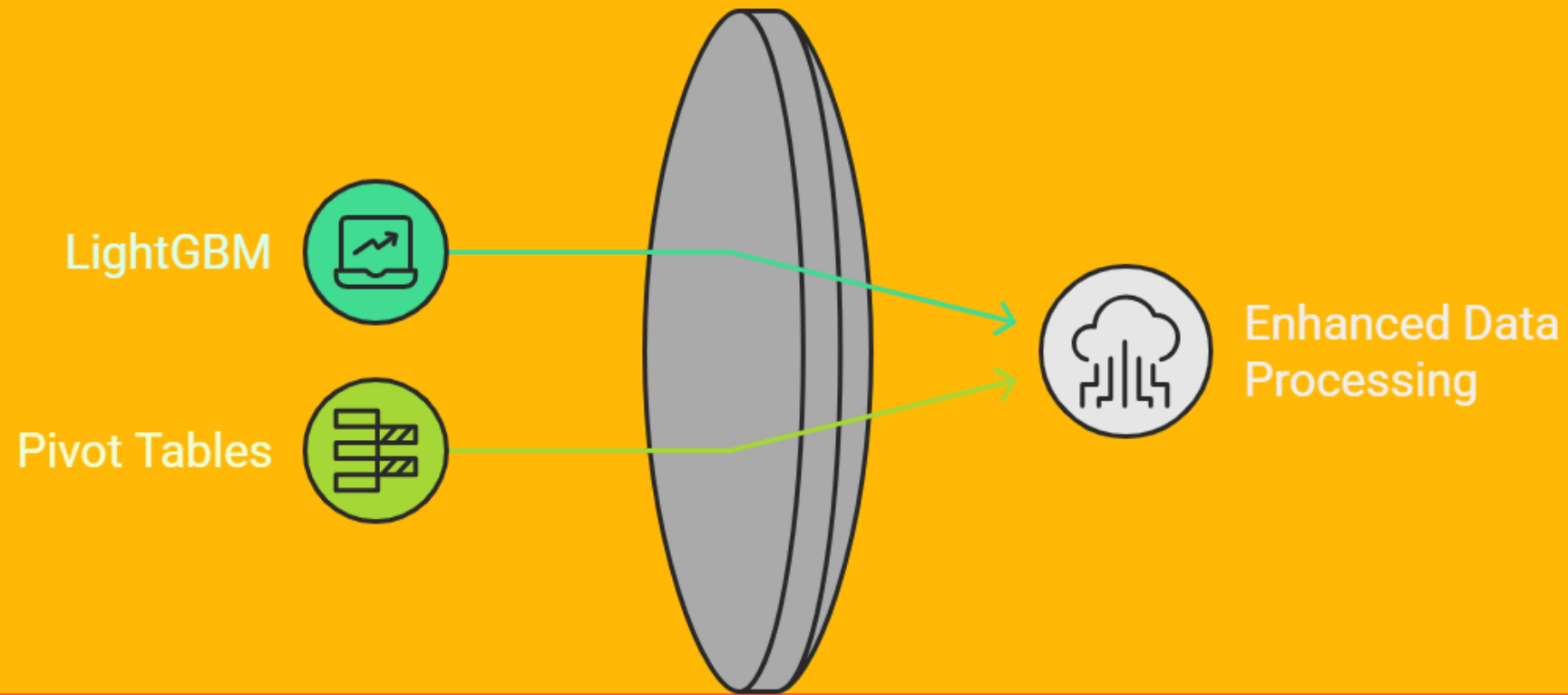


Clinical Trial Failure Rates

Research indicates that nearly 33% of clinical trials do not complete successfully due to various causes. This high attrition rate raises concerns regarding resource allocation and patient impact, necessitating a more predictive approach to trial management.

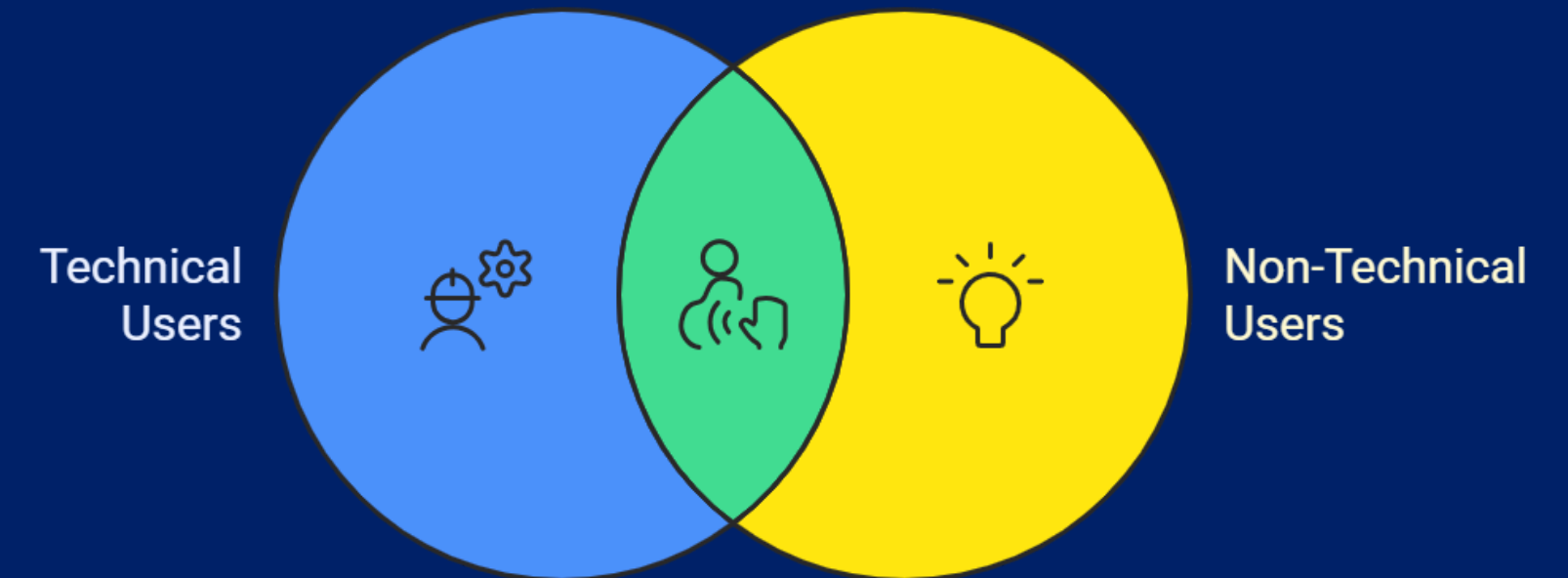
Unique Value Proposition

Advance Feature Selection Techniques

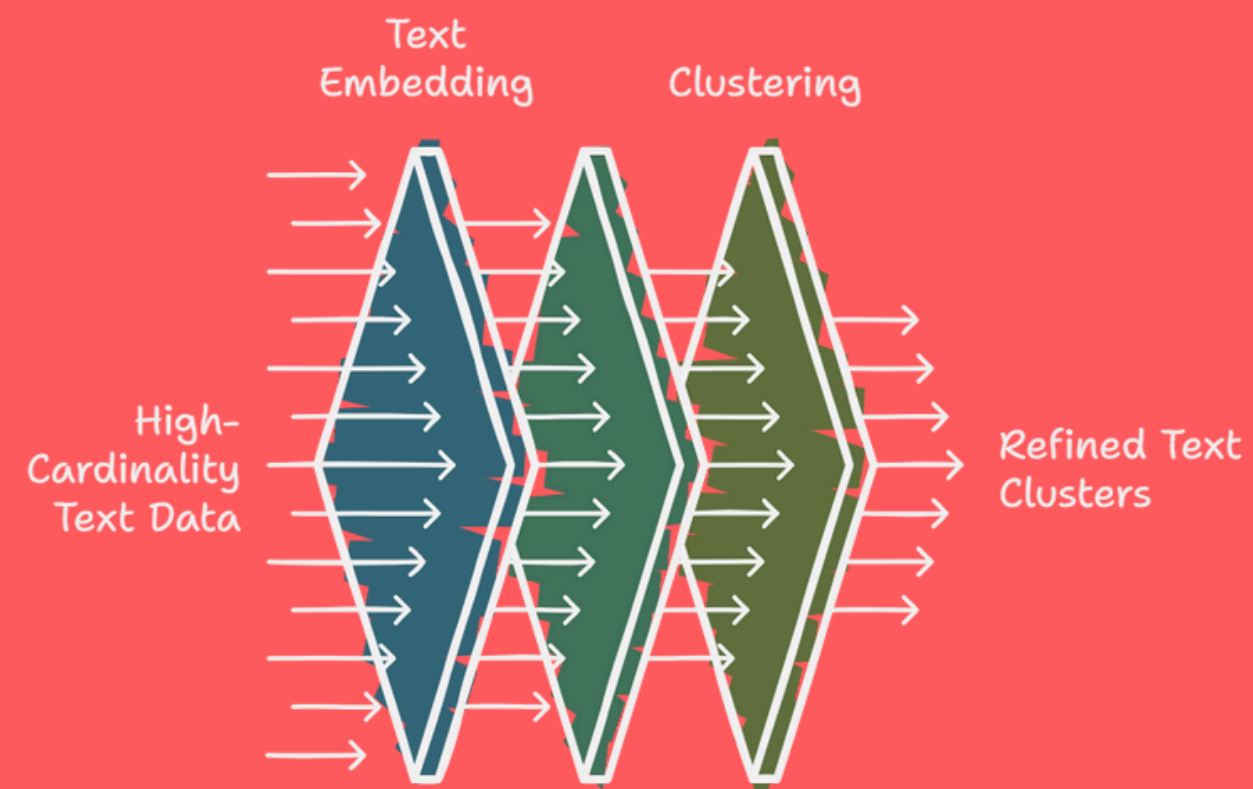


Inclusive Demo Design

Accessible and Interactive Interface



Handling high-cardinality text using NLP (Transformer + UMAP + HDBSCAN)



XGBoost Model Performance Metrics



Key Limitations and Challenges

Key limitations

Computational Constraints

Using Deep learning models for training was infeasible on local machines.

Feature Extraction

Generated unnecessary features; will use topic modeling for better feature extraction in the future.



- **High-Cardinality Features:** Required advanced encoding techniques to manage complexity.
- **Imbalanced Data:** Applied class weighting to ensure fair representation.
- **One-to-Many Relationships:** Carefully normalized and aggregated data for consistency.



1. Implement Topic Modeling

Steps:

- 1 Extract Features: Generate BioBERT embeddings for textual columns.**
- 2 Apply Topic Modeling: Use LDA (Latent Dirichlet Allocation) to extract 25 meaningful topics.**
- 3 Encode Each Row: Assign topic probabilities to each row based on its BioBERT embeddings.**

Next Steps

2. Enhance User Interface

Improve the web app UI/UX so that both technical & non-technical users can easily interact with results.

3. Deploy Model as an API-Based Solution

Convert the model into a Flask/FastAPI-based API for easy integration into web & enterprise applications.

