COMCAST

Call Me Maybe?

Predicting Churn in Real-time Using NLP & LLMs

Comcast Project Sponsors: Benjamin Land, Ian Tongs, & Jacob Linder
Faculty Advisor: James Butler
MBAn Students: Dilan SriDaran (dilan_s@mit.edu), Maxime Wolf (maximew@mit.edu)

PROBLEM STATEMENT

OBJECTIVE

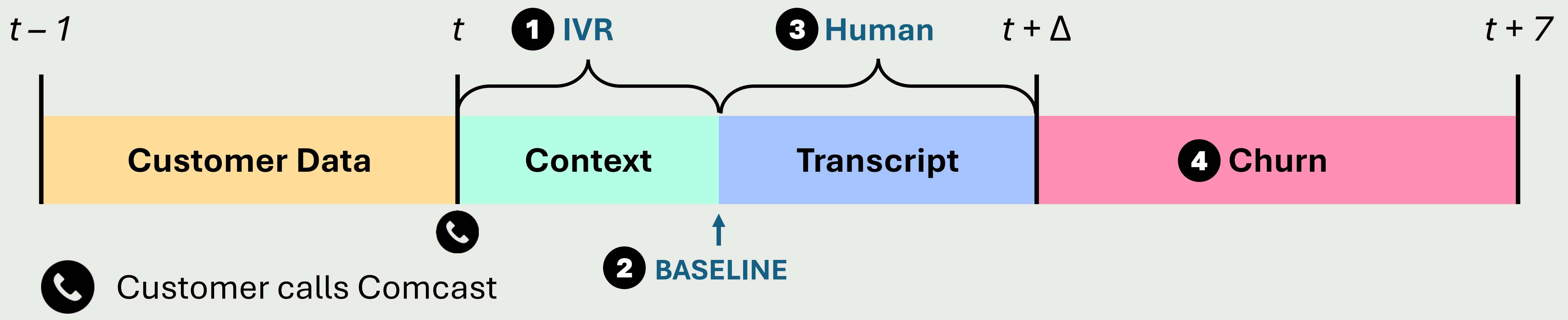
31M	**6M**	**$150**
Approx. Residential Customers	Approx. Monthly Call Volumes	Approx. Monthly Customer Revenue

Can transcripts give **better real-time offers to agents** and **more informative monitoring to management**?

DATA

Customer Data Demographic and account info	**Context** Customer intent from IVR
Transcript Digitized transcripts	**Churn** 7-day voluntary churn flag

CURRENT APPROACH



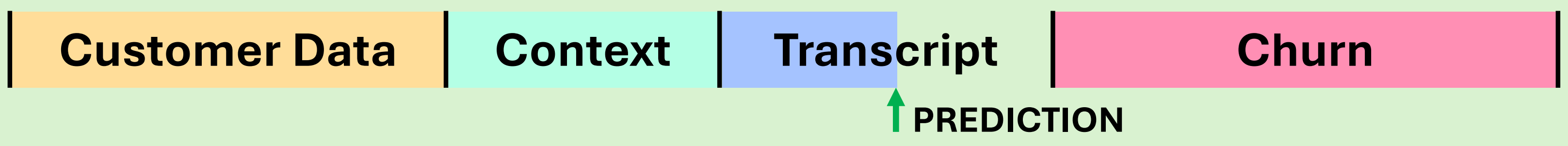
- ☎ Customer calls Comcast
- ① Interactive Voice Response (IVR) system identifies **customer's intent** ("context code")
- ② Existing "baseline" model predicts churn; automatically **suggests static offers** to agent
- ③ Human agent manages interaction; transcripts are digitized
- ④ 7-day voluntary churn (representing **savable** and **influenceable** customers) is monitored

IMPACT

REAL-TIME MODEL

IMPLEMENTED AFTER 100 WORDS OF TEXT

+10%	**$0.8M**
High-risk customers identified	Potentially savable monthly revenue

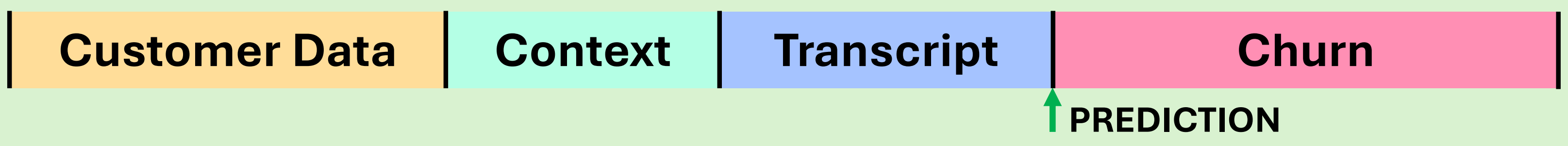


- ✓ Can provide **early, actionable, and dynamic insights** and offers
- ✗ Latency and model **simplicity is a key constraint** for deployment
- ✗ **Lower performance** due to less information from transcripts

NON-REAL-TIME MODEL

IMPLEMENTED AFTER CALL IS COMPLETE

+22%	**$1.7M**
High-risk customers identified	Potentially savable monthly revenue



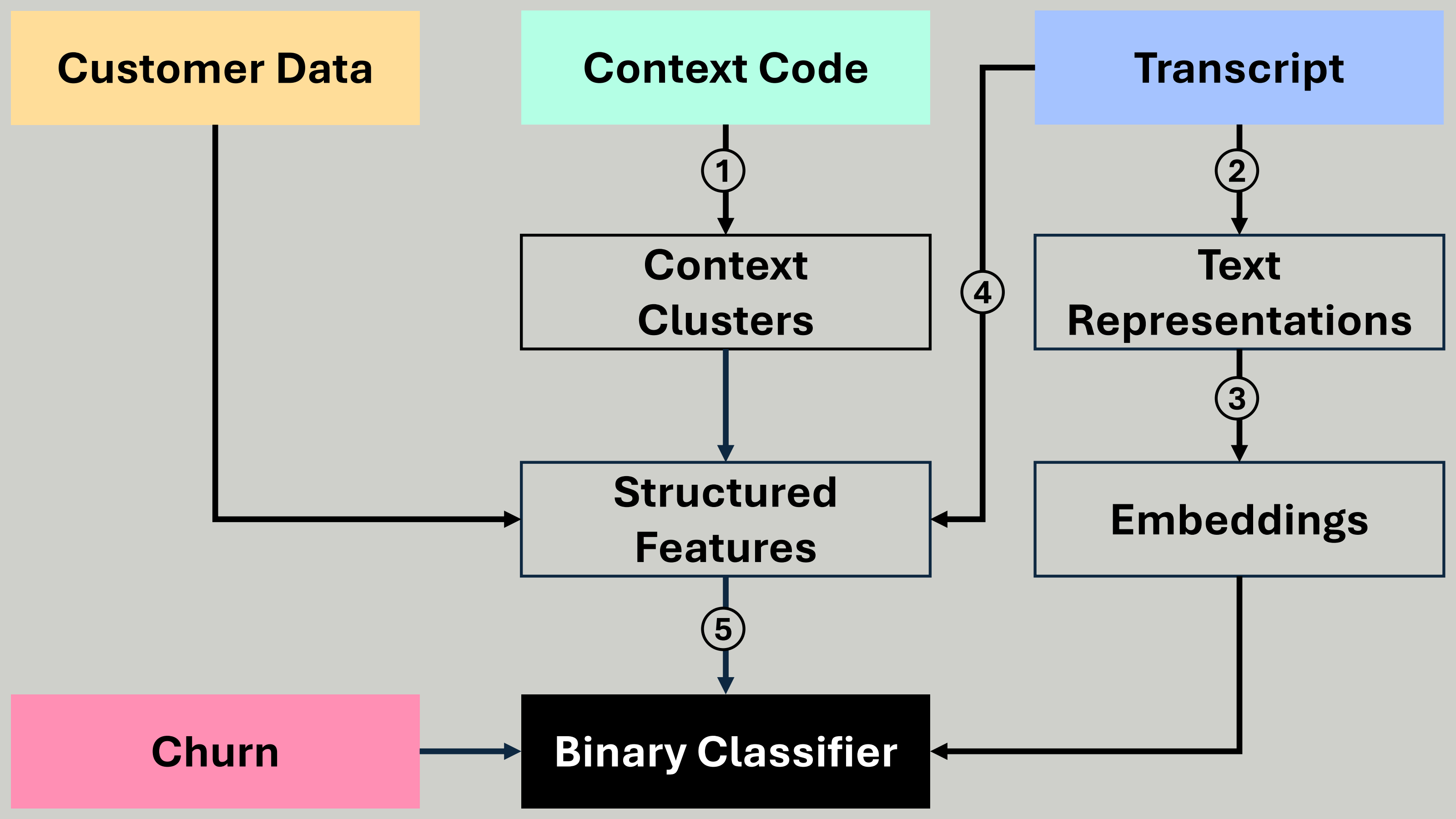
- ✓ Makes use of **all available information** from transcripts
- ✓ Latency and model **complexity is not a constraint**
- ✗ **Less scope to act** on predictions after a call

For internal security, results shown based on rounded assumptions and figures extracted from publicly released quarterly financials

APPROACH

GENERAL FRAMEWORK

We built a **multi-modal framework** that can **generalize to any binary target**, through either a non-real-time or real-time lens



① **Context Clusters**
What? Group 1,000+ codes into a small number of clusters to reduce feature granularity
How? kMeans clustering (with 50 clusters) applied to context code description embeddings

② **Text Representations**
What? Convert multiple utterances into single text string to represent transcript
How? Complete, Segments, and **LLM Summaries** (see section below)

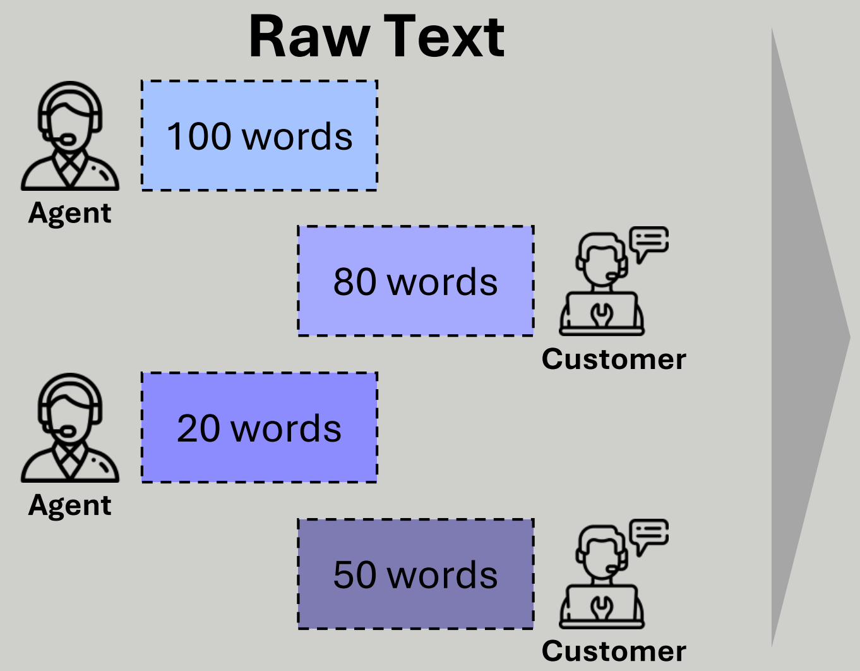
③ **Embeddings**
What? Convert single text strings into numerical vector representations
How? State-of-the-art large embedding model (**mxbai-embed-large**)

④ **Structured Features**
What? Extract interpretable structured features (e.g., competitor mentions) from transcripts
How? Mix of **business rules** and language models (**VADER**)

⑤ **Binary Classifier**
What? Train classification model using embeddings and features to predict 7-day churn
How? XGBoost (random search) with various other supervised learning algorithms tested

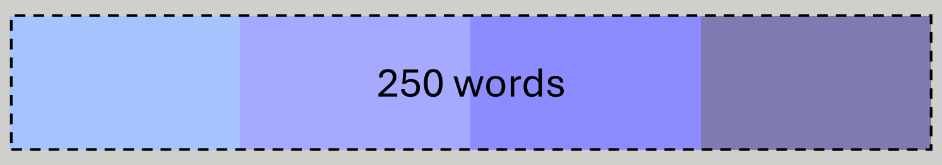
TEXT REPRESENTATIONS

Raw Text



Option 1: Complete

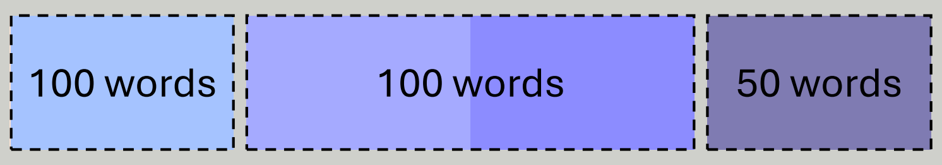
Single text string.
Single embedding.



- ✓ Computationally **lightweight**
- ✗ Embeddings often **poor on long texts**
- ✗ **Not applicable** for real-time settings

Option 2: Segments

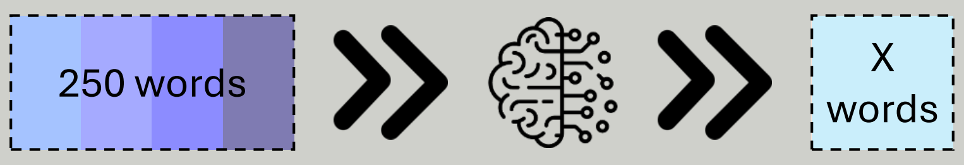
Multiple text strings. One embedding per string.
Average of embeddings up to current segment.



- ✓ Computationally **lightweight**
- ✓ Embeddings **better on shorter texts**
- ✓ **Suitable in both** real- and non-real-time

Option 3: LLM Summaries

Single text summary via LLM.
Single embedding.



- ✗ Summaries computationally **intensive**
- ✓ Embeddings **better on shorter texts**
- ✓ Enhances model **explainability**

RESULTS

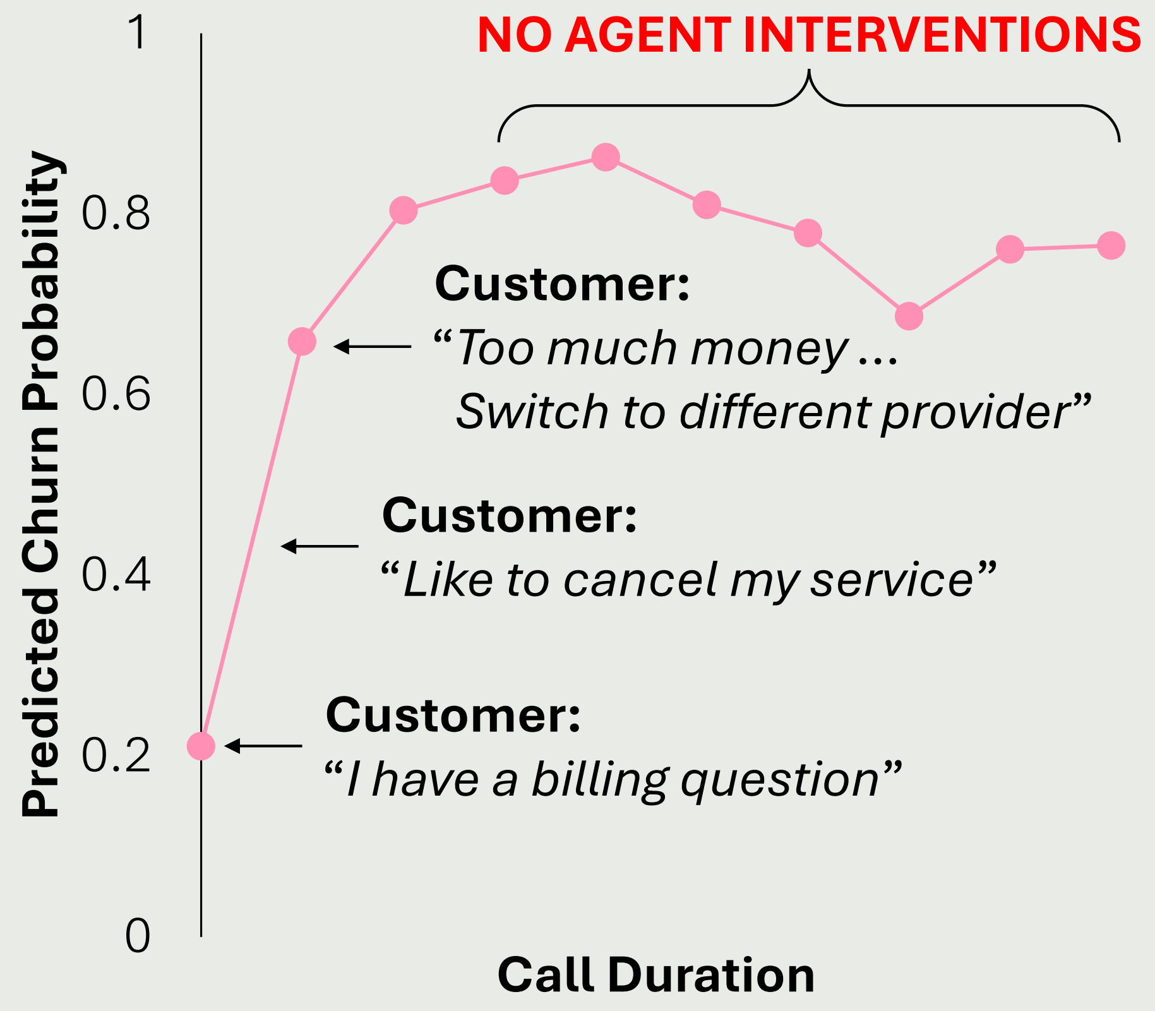
PERFORMANCE

Model	KS Score	KS Uplift to Base
BASELINE	**43.8**	
NON-REAL-TIME:		
Complete	55.7	+27%
Segments	**64.9**	**+48%**
LLM Summaries	59.1	+35%
REAL-TIME:		
Complete	*Not applicable*	
Segments	**61.9**	**+42%**
LLM Summaries	*Not feasible*	

Kolmogorov-Smirnov (KS) score measures purity of separation between predicted classes; higher value is better

KEY INSIGHTS

Real-world Example of Customer Churn



USE CASES

Real-time Offers
Re-run risk prediction early in call and revise offers; lower churn for high-risk customers and lower dilution for low-risk

Post-call Follow-ups
Targeted follow-up calls to high-risk customers to ensure their issues were resolved and/or offer promotions

Transcript Mining
Mine transcripts using risk scores to understand churn drivers and competitive landscape

THE ROADMAP

Data Pipeline
Explore and resolve data issues and ensure technology capability for real-time deployment

Detailed Costing
Undertake costing including investments, LLM and compute costs, potential uplift, and dilution

Trial
Implement formal trial to explore effectiveness of programs and any unintended consequences