PROJECT PROPOSAL DOCUMENT

Project Title: AI-Powered Customer Retention & Support Intelligence System

(Multimodal Machine Learning using Structured + Unstructured Customer Feedback)

---

1. Executive Summary

Customer churn has a direct financial impact on telecom companies. Traditional churn models rely only on structured customer data (billing, tenure, service features), but they ignore valuable insights hidden in **customer feedback text**.

This project develops an **AI-Powered Customer Retention & Support Intelligence System** using **multimodal machine learning**, which combines:

- **Structured telecom customer attributes**
- **Unstructured customer feedback text**
- **Churn labels**

Using the **Telco Customer Churn with Realistic Customer Feedback** dataset, this project predicts churn more accurately and reveals customer sentiments, dissatisfaction patterns, and actionable insights.

The system helps organizations make data-driven decisions, proactively address customer issues, and reduce churn, improving overall customer experience and revenue stability.

---

2. Problem Statement

Telecom companies lose significant revenue due to customer churn. While structured CRM data captures customer history, billing patterns, and services used, it does not reflect **customer emotions, complaints, or satisfaction levels**.

Companies often store customer feedback (emails, support conversations, survey comments) but rarely integrate it into churn prediction due to complexity.

The problem is:

How can we combine structured data and unstructured feedback text to predict churn and understand why customers leave?

This project addresses this by building a **multimodal churn prediction system** that fuses numerical features with textual sentiment and feedback topics.

---

3. Objectives

1. Build a **predictive model** to identify customers likely to churn.
2. Analyze **customer feedback text** using NLP techniques.
3. Develop a **multimodal machine learning pipeline** that integrates structured and unstructured data.
4. Extract **key churn drivers** using explainable AI.
5. Build an **interactive dashboard** to visualize churn risk and customer sentiment.
6. Provide actionable insights for **customer retention strategies**.

---

4. Dataset Description

📌 **Dataset: *Telco Customer Churn with Realistic Customer Feedback***

🔗 https://www.kaggle.com/datasets/beatafaron/telco-customer-churn-realistic-customer-feedback

This dataset is specifically designed for **multimodal machine learning** and includes:

Structured Features

- **CustomerID**
- Gender
- SeniorCitizen
- Partner
- Dependents
- Tenure
- PhoneService
- MultipleLines
- InternetService
- OnlineSecurity
- DeviceProtection
- TechSupport
- StreamingTV
- Contract
- PaperlessBilling
- PaymentMethod

- MonthlyCharges
- TotalCharges

Unstructured Feature (Text)

- **Customer Feedback Text**
    - Realistic feedback generated for each customer
    - Includes complaints, praise, dissatisfaction, service issues
    - Examples:
        - "Internet speed is slow and customer support is unhelpful."
        - "Billing errors keep happening every month."
        - "Very satisfied with the service."

Target Variable

- **Churn (Yes/No)**

Combined Dataset Concept

This dataset already includes both modalities for each customer:

CustomerID	Tenure	Charges	Payment	Feedback Text	Churn
7590-VHVEG	5	75.65	Electronic	"Internet interruptions frequently…"	Yes
5575-GNVDE	2	53.85	Bank Transfer	"Service is smooth and reliable."	No

Thus, no synthetic merging is required.

---

5. Methodology – Step-by-Step Implementation Plan

Step 1 – Data Understanding

- Explore the structure, types, and distributions.
- Identify key variables influencing churn.

Step 2 – Data Preprocessing

- Handle missing values and convert TotalCharges into numeric.
- Encode categorical variables (Contract, PaymentMethod, etc.)
- Clean feedback text:

    - Remove special characters
    - Lowercase
    - Remove stopwords
    - Lemmatization

Step 3 – Exploratory Data Analysis (EDA)

- Churn distribution and imbalance check
- Monthly charges vs churn
- Contract type vs churn
- Sentiment distribution of feedback
- Word clouds for churned vs non-churned customers

Step 4 – NLP Processing

- Convert feedback text into features using:
    - TF-IDF Vectorizer, or
    - BERT Embeddings (advanced)
- Extract:
    - **Sentiment score**
    - **Polarity & subjectivity**
    - **Keyword topics**

Step 5 – Feature Engineering

- Combine:
    - Structured numerical/categorical features
    - Text-derived features (sentiment, embeddings)
- Create new features:
    - Complaint length
    - Sentiment category (positive/neutral/negative)
    - Tenure bands

Step 6 – Model Development

- Train multiple models:
    - Logistic Regression, Random Forest, XGBoost
    - Multimodal Neural Network (text + numeric fusion)
- Compare structured-only vs multimodal performance.

Step 7 – Model Evaluation

- Use metrics:
  - Precision / Recall
  - F1-score
  - ROC-AUC
- Compare performance of text-based model vs structured model.

Step 8 – Explainability

- Use SHAP to understand:
  - Top churn indicators
  - Impact of negative feedback
  - Contract & billing influences

Step 9 – Deployment

Deploy using:

- **Streamlit Web App**, or
- **Power BI Dashboard**

The app will:

- Accept customer details
- Accept feedback text
- Predict churn probability
- Show sentiment & key risk factors

---

6. Workflow Overview

Data Loading
↓
Preprocessing (Structured + Text Cleaning)
↓
Exploratory Data Analysis
↓
Sentiment & Text Embedding Extraction
↓
Feature Engineering & Fusion
↓
Model Training (Multimodal)
↓
Evaluation & Explainability
↓
Deployment (App / Dashboard)

7. Tools & Technologies

Category	Tools
Programming	Python
Data Processing	Pandas, NumPy
ML	Scikit-learn, XGBoost
NLP	NLTK, SpaCy, HuggingFace Transformers
Visualization	Matplotlib, Seaborn, WordCloud
Deployment	Streamlit, Flask, Power BI
Explainability	SHAP
Version Control	Git, GitHub

---

8. Hardware & Software Requirements

Hardware

- 8–16 GB RAM
- Intel i5/Ryzen 5 or above
- GPU optional (for BERT models)

Software

- Python 3.10+
- Jupyter Notebook
- Streamlit or Power BI Desktop (optional)
- Required libraries installed through pip/anaconda

---

9. Expected Outcomes

- Accurate churn prediction with multimodal data
- Sentiment insights from customer feedback text
- Identification of top churn drivers
- Dashboard showing:
    - Churn probability, Sentiment score, Customer complaints analysis
- A real-world, industry-standard project

---

10. Business Impact

This AI system helps telecom companies:

- **Reduce churn proactively**
- Improve customer satisfaction
- Prioritize at-risk customers
- Improve support operations
- Allocate retention budgets more effectively

This directly improves **revenue stability & customer lifetime value (CLTV)**.

---

11. Challenges & Limitations

- Some sentiment text may be ambiguous
- Models may require hyperparameter tuning
- Dataset may still have class imbalance
- BERT models require more compute resources

---

12. Future Enhancements

- Add voice sentiment analysis (call center audio)
- Integrate real-time streaming feedback
- Add customer segmentation (clustering)
- Build automated retention recommendations using LLMs
- Deploy on cloud (AWS / GCP / Azure)

---

13. Conclusion

This project provides a **complete AI-driven solution** for customer retention by integrating structured CRM data with rich customer feedback text.
It demonstrates advanced skills in:

- Machine learning, NLP
- Multimodal data fusion
- Customer analytics
- Model deployment
- Explainable AI

This is exactly the kind of project that **top companies look for in Data Science portfolios**.

14. Deliverables

1. Final Multimodal Dataset
2. Data Preprocessing Report
3. EDA Report (with visualizations)
4. Text Sentiment Analysis Report
5. ML Model & Evaluation Report
6. Streamlit App / Power BI Dashboard
7. Final Project Report (this proposal)
8. GitHub Repository with documentation

🌟 Final Project Summary

The **AI-Powered Customer Retention & Support Intelligence System** uses multimodal machine learning to predict churn and analyze customer feedback using structured CRM features and realistic text data.
It delivers actionable insights, improves decision-making, and showcases strong technical and business problem-solving skills — making it an exceptional project for job portfolios, hiring managers, and real-world telecom analytics.