Churn Prediction and Retention System - Project Plan

Project Roadmap
This project involves multiple stages: **data preprocessing, predictive modeling, NLP analysis, Generative AI for ret

---

Step-by-Step Plan

1. Data Preparation & MySQL Setup
- Upload the dataset (`customer churn data_usecase2_Hackathon.xlsx`) to **MySQL**:
- Convert the Excel sheet into a **structured relational database**.
- Create a **table schema** matching the dataset features.
- Use **SQLAlchemy** in Colab to fetch data from MySQL.

- **Understanding Dataset**:
- Load data from MySQL into **Pandas DataFrame**.
- Check **missing values, data types, and outliers**.
- Identify **target variable** (`churned/not churned`).
- Perform **exploratory data analysis (EDA)**:
- **Visualizations** (histograms, box plots, correlations).
- **Feature importance ranking**.

---

2. Predictive Modeling for Churn
- Select at least **two machine learning models**:
- Logistic Regression (Baseline).
- Random Forest / XGBoost (for better performance).

- **Feature Engineering**:
- Convert categorical variables using **One-Hot Encoding**.
- Handle **imbalanced data** (SMOTE / oversampling).
- Use **feature selection techniques** (SHAP values, mutual information).

- **Model Training & Evaluation**:
- Train models on historical churn data.
- Compare **accuracy, precision-recall, F1-score, AUC-ROC**.
- Use **Hyperparameter tuning** for optimization.

- **Insights**:
- Identify **top factors contributing to churn**.

---

3. NLP for Customer Feedback Analysis
- **Data Preprocessing**:
- Convert textual feedback into **structured data**.
- Remove **stop words, lemmatization, stemming**.

- **Sentiment Analysis**:
- Use **VADER** or **TextBlob** for sentiment scoring.
- Identify **common customer complaints**.

- **Topic Modeling**:
- Apply **LDA (Latent Dirichlet Allocation)** to group customer concerns.

- **Key Insights Extraction**:
- Map sentiment trends to **churn behavior**.

---

4. Generative AI for Retention Strategies
- **Develop Personalized Retention Strategies**:
- Use insights from churn models and NLP analysis.
- Develop **targeted retention plans** (discounts, loyalty programs).

- **Chatbot Integration**:
- Build a **Generative AI chatbot** using **LangChain + Hugging Face**.
- Implement **multi-turn conversation** for handling customer complaints.

- **Example Chatbot Logic**:
- **User Query**: *"My card keeps getting blocked"*
- **Response**: *"I understand your issue. Would you like me to escalate this to the Debit Card team?"*

- **Evaluation & Refinement**:
- Assess chatbot responses using **human feedback**.
- Implement **automated learning from interactions**.

---

5. System Integration & Deployment
- **Architecture Design**:
- **MySQL for data storage**.
- **Colab for model training & evaluation**.
- **FastAPI for serving predictions & chatbot interactions**.
- **Streamlit / Flask for UI (if needed for visualization)**.

- **Deployment Plan**:
- Choose **Google Cloud / AWS for cloud-based deployment**.
- Implement **real-time data updates**.

- **Scalability Considerations**:
- Optimize **database queries & indexing**.
- Use **batch processing for large data handling**.

---

6. Ethical Considerations & Privacy
- **Fairness & Bias Mitigation**:
- Ensure models **don't discriminate based on sensitive attributes**.
- Regularly audit **feature importance**.

- **Customer Privacy**:
- Implement **data encryption** and **GDPR compliance**.

- **Transparent AI**:
- Provide **explainability reports** for stakeholders.

Deliverables
- **Colab Notebooks**: End-to-end implementation.
- **Python Scripts**: For final model deployment.
- **Presentation (10 Slides)**: Summary, findings, recommendations.
- **Requirements.txt**: List of libraries used.

Next Steps
1■■ **Upload data to MySQL and verify schema.**
2■■ **Connect MySQL to Google Colab & perform EDA.**
3■■ **Train churn prediction models and evaluate.**
4■■ **Implement NLP for feedback analysis.**
5■■ **Develop chatbot and test retention strategies.**
6■■ **Integrate all components and deploy.**