Customer Churn Prediction Model

1. Executive Summary

Problem Statement: Your business faces a challenge in identifying customers at risk of leaving. Without a proactive strategy, you lose valuable customers, impacting revenue and growth.

Proposed Solution: We propose developing a machine learning model that predicts customer churn probability. This AI model will analyze customer behavior and demographic data to provide an accurate risk score for each customer, enabling your teams to intervene with targeted retention campaigns.

Expected Business Impact: This project will help you reduce customer churn by an estimated 10-15%, leading to increased customer lifetime value and significant revenue savings.

2. Introduction and Background

In today's competitive market, retaining existing customers is far more cost-effective than acquiring new ones. Currently, your business relies on manual or reactive methods to identify at-risk customers, which often come too late. This project will implement a data-driven, predictive approach to identify and address churn before it occurs.

3. Project Scope and Deliverables

Detailed Scope:

Data Collection & Preparation: We will work with your team to collect and clean historical customer data, including transaction history, customer service interactions, and demographic information.

Feature Engineering: We will create new variables from raw data to improve model performance, such as "days since last purchase" or "average monthly spend."

Model Training & Validation: We will train several machine learning models (e.g., Logistic Regression, Gradient Boosting) and select the best-performing one based on your business needs. The model will be validated on a separate dataset to ensure its accuracy.

API Development: The final model will be deployed as a secure API endpoint, allowing your CRM or marketing automation system to query a customer's churn risk score in real time.

Dashboard: We will provide a simple dashboard to visualize churn trends and monitor model performance.

Non-Deliverables: This project does not include implementing new marketing campaigns or customer service interventions based on the model's output. It also does not cover real-time streaming data ingestion.

4. Technical Approach and Methodology

AI/ML Technique: We will use supervised machine learning, specifically classification algorithms, to predict a binary outcome (churn or not churn). We will explore models like XGBoost or Random Forest for their high performance and interpretability.

Development Lifecycle: The project will follow an agile methodology with three key phases:

Data Ingestion & Exploration (4 weeks): Connect to your data sources, clean the data, and perform exploratory analysis.

Model Development & Training (6 weeks): Engineer features, train and tune various models, and select the final model.

Deployment & Integration (4 weeks): Package the model as an API, set up the production environment, and assist with integration testing.

Technology Stack: Python, Pandas, Scikit-learn, XGBoost, Flask/FastAPI, Docker, and a cloud platform like AWS or Azure.

Model Performance Metrics: The model will be evaluated based on Precision, Recall, and the F1-score to ensure it accurately identifies at-risk customers without creating too many false positives.

5. Project Timeline and Milestones

Timeline: 3-4 months

Key Milestones:

Month 1: Data Access and Initial Analysis.

Month 2: Feature Engineering and Model Training.

Month 3: Model Validation and API Development.

Month 4: Final Integration and Documentation.

6. Team and Resources

Team Roles: 2 Data Scientists/ML Engineers.

Man-days: Approximately 180-240 man-days.

Client Responsibilities: Provide access to historical customer data, dedicate a point of contact for project collaboration, and allocate internal resources for system integration.

7. Pricing and Payment Schedule

Total Project Cost: $50,000 - $80,000

Cost Breakdown:

Data Scientist Time: 60%

Infrastructure/Software: 15%

Project Management & Overhead: 25%

Payment Terms: 30% upfront, 30% after model validation, 40% upon final deployment.

8. Risk Assessment and Mitigation

Risk: Poor data quality or insufficient historical data.

Mitigation: We will begin with a data quality assessment. If data is an issue, we will work with your team to determine if additional data sources are available or if we need to adjust the project scope.

Proposal 2: Predictive Maintenance for Industrial Equipment

1. Executive Summary

Problem Statement: Your manufacturing operations are subject to unplanned equipment downtime, leading to production delays, increased maintenance costs, and potential safety risks.

Proposed Solution: We propose building an AI model that predicts equipment failures before they happen. The model will analyze real-time sensor data from machinery to identify patterns that precede a breakdown, allowing for scheduled, proactive maintenance.

Expected Business Impact: This solution will reduce unplanned downtime by over 30%, extend equipment lifespan, and lower maintenance costs by optimizing repair schedules.

2. Introduction and Background

Traditional, time-based maintenance often results in either performing maintenance too early or too late. By leveraging machine learning and IoT sensor data, we can shift to a condition-based maintenance strategy, making your operations more efficient and reliable.

3. Project Scope and Deliverables

Detailed Scope:

Sensor Data Ingestion: Develop a system to ingest real-time sensor data (e.g., temperature, vibration, pressure, current) from up to 100 pieces of equipment.

Anomaly Detection Model: Train a deep learning model to recognize normal operating conditions and flag any deviations that indicate a potential failure.

Failure Prediction Model: Develop a separate model that, based on historical failure data, predicts the time-to-failure for critical components.

Real-time Dashboard: Create a dashboard that displays the health status of each piece of equipment, potential failure alerts, and maintenance recommendations.

Alert System: Integrate an alert system to send notifications via email or SMS to maintenance crews when a high-risk event is detected.

Non-Deliverables: We will not be responsible for installing new hardware sensors or for the physical maintenance of the equipment itself.

4. Technical Approach and Methodology

AI/ML Technique: We will use time-series analysis and deep learning (e.g., LSTM or Transformer models) to handle the sequential nature of sensor data. Anomaly detection will use techniques like Isolation Forest or Autoencoders.

Development Lifecycle:

Data & Infrastructure Setup (6 weeks): Configure data pipelines for real-time sensor data and set up a robust cloud infrastructure.

Model Training & Backtesting (10 weeks): Train the anomaly and prediction models on historical data and backtest their performance.

Deployment & Pilot (8 weeks): Deploy the models and dashboard, and run a pilot program on a select number of machines to validate the system.

Technology Stack: Python, TensorFlow/PyTorch, Apache Kafka for data streaming, Grafana for the dashboard, and a cloud environment (e.g., AWS IoT, Google Cloud).

Model Performance Metrics: The model will be evaluated on its Root Mean Squared Error (RMSE) for time-to-failure predictions and Precision/Recall for anomaly alerts.

5. Project Timeline and Milestones

Timeline: 5-7 months

Key Milestones:

Month 2: Sensor Data Pipeline Operational.

Month 4: Anomaly Detection Model Ready for Testing.

Month 6: Predictive Maintenance System Deployed for Pilot.

6. Team and Resources

Team Roles: 3-4 Data Scientists/ML Engineers, 1 Back-end Developer.

Man-days: Approximately 450-700 man-days.

Client Responsibilities: Provide access to sensor data, share historical failure logs, and identify key personnel from maintenance and IT to collaborate on the project.

7. Pricing and Payment Schedule

Total Project Cost: $90,000 - $150,000

Cost Breakdown:

ML & Development Time: 65%

Infrastructure & Software: 20%

Project Management & Overhead: 15%

Payment Terms: 25% upfront, 25% at infrastructure setup completion, 25% after successful model backtesting, and 25% upon final deployment.

8. Risk Assessment and Mitigation

Risk: Insufficient or noisy sensor data.

Mitigation: We will conduct an initial data quality and availability check. If data is incomplete, we may need to adjust the project scope or timeline to account for data collection efforts.