# Agent Churn Prediction Model Documentation

## 1. Introduction

This document provides a detailed overview of the agent churn prediction model developed using various machine learning algorithms. The model was trained and evaluated on a dataset to predict churn for agents. This document includes descriptions of the algorithms used, their performance metrics, and details about the final selected model.

## 2. Dataset

**Agent Master Data**

Description: The dataset contains information about agents, including derived features relevant for predicting churn.

## 3. Data Preprocessing

**3.1. Feature and Target Variables**

Features: All columns except Agent ID and Churn.

Target Variable: Churn, churn probability

**3.2. Preprocessing Steps**

Numerical Data:

- Imputation of missing values using median.

- Standardization using StandardScaler.

Categorical Data:

- Imputation of missing values using the most frequent value.

- Encoding using OneHotEncoder.

## 4. Machine Learning Models

### 4.1. Logistic Regression

Description: A statistical model that uses a logistic function to model a binary dependent variable.

**Evaluation Metrics:**

| **Metric** | **Value** |
| --- | --- |
| Accuracy | 0.8545 |
| Precision | 0.7972 |
| Recall | 0.7655 |
| F1 Score | 0.7810 |
| Confusion Matrix | [[1190 132]  [ 159 519]] |
|  |  |

### 4.2. Gradient Boosting Classifier

Description: An ensemble method that builds models sequentially to correct errors made by previous models.

**Evaluation Metrics:**

| **Metric** | **Value** |
| --- | --- |
| Accuracy | 0.9775 |
| Precision | 0.9817 |
| Recall | 0.9513 |
| F1 Score | 0.9663 |
| Confusion Matrix | [[1310 12]  [ 33 645]] |

### 4.3. Support Vector Classifier (SVC)

Description: A classification method that finds the hyperplane which best separates the classes in the feature space.

**Evaluation Metrics:**

| **Metric** | **Value** |
| --- | --- |
| Accuracy | 0.8845 |
| Precision | 0.8498 |
| Recall | 0.8009 |
| F1 Score | 0.8246 |
| Confusion Matrix | [[1226 96]  [ 135 543]] |
|  |  |

## 5. Final Model Selection

After evaluating the models, the Gradient Boosting Classifier was selected as the final model due to its superior performance across all evaluation metrics.

**Final Model Details**

Model Used: Gradient Boosting Classifier

**Evaluation Metrics on Test Data:**

- Accuracy: 0.9775

- Precision: 0.9817

- Recall: 0.9513

- F1 Score: 0.9663

## 6. Model Predictions

**6.1. Test Data Predictions**

Predictions were made on the test dataset and saved to a CSV file:

**6.2. All Data Predictions**

Predictions and churn probabilities were generated for all rows in the dataset and saved to a CSV file:

**File Path:** C:/Users/Administrator/Desktop/Internship work/Agent Churn Prediction Model/Agent\_Churn\_Predictions\_with\_Probabilities.csv

**Columns in the file:**

- Agent ID

- Churn Label (Predicted Churn)

- Churn Probability (Probability of Churn)

## 7. Conclusion

The Gradient Boosting Classifier model demonstrates high accuracy and robust performance for predicting agent churn. The generated predictions and probabilities provide valuable insights into churn likelihood, which can be used for further analysis and decision-making.