**Customer Churn Prediction Report**

**1. Introduction**

Customer churn prediction plays a vital role in retaining valuable customers and reducing revenue loss. In this project, we developed a machine learning model to accurately predict whether a customer is likely to churn or not. This model can assist the business in taking timely actions to retain high-risk customers and improve overall customer satisfaction.

**2. Selected Algorithm and Rationale**

Algorithm: Random Forest Classifier

The Random Forest Classifier was selected for this task based on the following reasons:

* Handles class imbalance well, which was a key challenge in our dataset (fewer churned customers than non-churned).
* It is a robust ensemble method that reduces overfitting by combining multiple decision trees.
* Performs well on both classification and regression tasks and is known for high accuracy and interpretability.
* It automatically handles feature interactions and non-linear patterns without requiring manual feature engineering.

**Initially, Logistic Regression was tested but it failed to correctly predict churned customers due to class imbalance. Random Forest significantly improved the performance.**

**3. Model Training and Evaluation**

**Model Setup:**

* Data was preprocessed using one-hot encoding.
* The dataset was split into training and testing sets using an 80:20 ratio.
* The Random Forest model was trained with 100 estimators (n\_estimators=100) and a fixed random seed for reproducibility.

**Evaluation Metrics:**

|  |  |
| --- | --- |
| **METRIC** | **VALUE** |
| **Accuracy** | **98%** |
| **Precision(class-1 churned)** | **100%** |
| **F1 score(class-1 churned)** | **0.97** |
| **Confusion Matrix** | **[[826, 0], [13, 202]]** |

**Interpretation:**

* The model correctly predicted 202 out of 215 churned customers.
* It also correctly identified all non-churned customers (826/826).
* The F1-score of **0.97** for churn class indicates **excellent detection of risky customers**.
* Compared to Logistic Regression (which had 0% recall on churn class), Random Forest showed major improvement.

**4. Business Utilisation and Recommendations**

**Utilising the Model's Predictions:**

* The business can **target customers predicted to churn** with retention strategies such as:
  + Personalized offers and discounts.
  + Loyalty programs or proactive service follow-ups.
  + Feedback requests to address dissatisfaction.
* High-risk customer segments can be prioritized by the marketing team for **retention campaigns**.
* Integration of model predictions with CRM tools can help **automate alerts** and trigger timely actions.

**Potential Areas for Improvement:**

* **Hyperparameter Tuning**: Use GridSearchCV to further optimize the model.
* **Feature Engineering**: Add more behaviour-based features (e.g., last login date, complaint frequency).

**5. Conclusion**

The implemented Random Forest model successfully predicts customer churn with high accuracy and recall. It enables the business to proactively intervene and retain valuable customers. This predictive system, when integrated with business strategy, can **significantly reduce churn rates** and improve customer satisfaction.