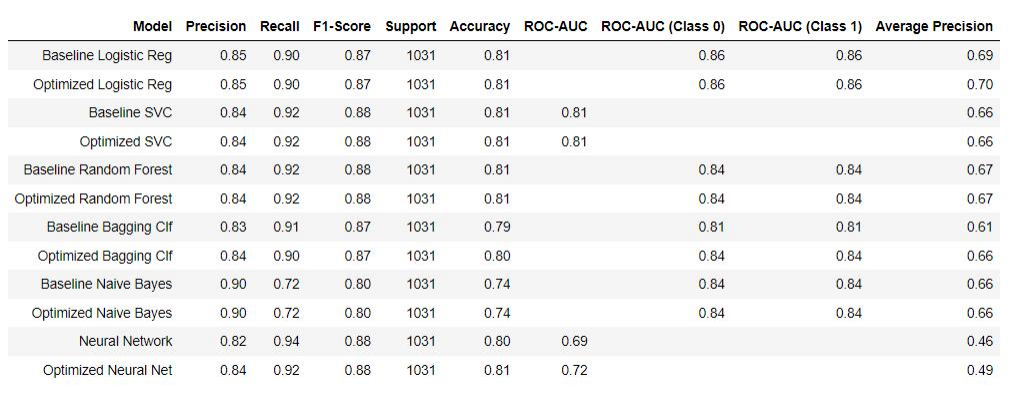
**Documentation Report: Customer Churn Prediction**

**Overview**

This report outlines the development , evaluation and deployment of a Support Vector Classifier (SVC) model for predicting customer churn which was shortlisted from the modelling process from phase 2 (candidate models listed below in figure). The focus was on hyperparameter optimization and regularization to enhance the model's performance.



**Hyperparameter Tuning**

**Methodology**

* **Approach**: Grid search was employed for hyperparameter tuning of the SVC model.
* **Parameters Considered**:
  + **C (Regularization Parameter)**: Controls the trade-off between increasing the margin size and ensuring that the samples are classified correctly.
  + **Kernel**: Transforms the feature space. Options tested included 'linear', 'rbf', 'poly', and 'sigmoid'.
* **Validation Method**: 5-fold cross-validation was used to evaluate the performance of different hyperparameter combinations.

**Results**

* **Optimal Hyperparameters**:
  + **C**: 1
  + **Kernel**: 'rbf'

**Model Training**

* **Configuration**: The SVC model was trained with C=1 and Kernel='rbf'.
* **Data Used**: Training was conducted using the provided training dataset (X\_train, y\_train).

**Model Effectiveness**

**Performance Metrics**

* **Accuracy**: Achieved an accuracy of 81% (0.81), indicating the proportion of correct predictions.
* **Precision and Recall**:
  + **Class 0**:
    - Precision: 0.84 (Model is correct 84% of the time when predicting class 0).
    - Recall: 0.92 (High ability to recognize true positives for class 0).
  + **Class 1**:
    - Precision: 0.69 (Lower reliability for class 1 predictions).
    - Recall: 0.52 (Less effective in recognizing true positives for class 1).
* **F1-Score**:
  + Class 0: 0.88
  + Class 1: 0.60
* **ROC-AUC**: The model scored 0.81, indicating good discriminatory power between classes.
* **Average Precision**: 0.66, reflecting the precision-recall trade-off.

**Insights**

* The hyperparameter tuning significantly improved the SVC model's performance, particularly with the RBF kernel and a regularization value of C=1.
* A notable performance disparity was observed between class 0 and class 1 predictions, with class 0 predictions being more accurate.
* The model demonstrates strong predictive power but requires further refinement for class 1 predictions.

**Recommendations**

* **Model Improvement**:
  + Investigate alternative kernels or regularization values to enhance class 1 predictions.
  + Consider additional features or data preprocessing techniques to balance the model's performance across classes.
* **Further Analysis**:
  + Examine the characteristics of class 1 to understand why the model is less effective for these predictions.
* **Model Evaluation**:
  + Additional metrics, such as a confusion matrix, could provide deeper insights into the types of errors made by the model.
* **Deployment Considerations**:
  + Before deploying the model, it's important to ensure it performs well across various customer segments and operational scenarios.

## Streamlit Web Application Functionality

### Overview

The Streamlit web application is an interactive interface designed for end-users to utilize the SVC model for customer churn prediction. It offers two primary functionalities:

1. **Single Customer Prediction**
2. **Batch Prediction via CSV Upload**

### Single Customer Prediction

#### Functionality

* Users can input details of a single customer into the web application.
* The input fields correspond to the features used by the SVC model.
* Once the details are entered, the user can submit the data for prediction.

#### Process Flow

1. **Input Data**: The user fills in the customer's details in the provided input fields.
2. **Submit for Prediction**: Upon submission, the data is processed by the SVC model.
3. **Display Results**: The application displays the churn prediction result, indicating whether the customer is likely to churn or not.

### Batch Prediction via CSV Upload

#### Functionality

* Users have the option to upload a CSV file containing details of multiple customers.
* The CSV file should follow the format and structure expected by the model, with columns corresponding to customer features.

#### Process Flow

1. **CSV Upload**: The user uploads a CSV file with the batch of customer data.
2. **Data Processing**: The application processes the entire batch through the SVC model.
3. **Visualization and Results**:
   * A pie chart visualization is generated, showing the proportion of customers likely to churn versus those who are not.
   * Alongside the pie chart, a detailed prediction result for each customer is displayed, categorizing them as 'Churn: Yes' or 'Churn: No'.

### User Interface

* The Streamlit web application is designed to be user-friendly and intuitive.
* Clear instructions and labels are provided for each input field and functionality.
* Error handling and user feedback mechanisms are integrated to guide users in case of incorrect data input or file format issues.

### Technical Considerations

* **Data Handling**: Proper validation and preprocessing are applied to user inputs to align with the model's requirements.
* **Security and Privacy**: The application ensures user data privacy and security, with no storage of personal or sensitive data.
* **Performance**: Optimized for efficient processing, even for batch predictions, to provide quick and accurate results.

## Reccomendations based on problem statement

**Targeted Intervention Strategies:** With an accuracy of 81%, users can reliably use the model to identify customers at high risk of churn. Businesses can develop targeted intervention strategies for these customers, such as personalized offers, improved customer service, or addressing specific concerns that might be leading to churn.

**Resource Optimization:** By accurately predicting which customers are likely to churn, companies can allocate their resources more efficiently. Instead of spreading resources thinly across all customers, they can focus on those identified by the model, leading to more cost-effective retention strategies.

**Customer Segmentation for Tailored Communication:** The model's precision and recall metrics for different classes suggest that it can be used for customer segmentation. Companies can tailor their communication and retention strategies based on the model’s classification, offering more personalized and relevant interactions.

**Performance Improvement for Class 1 Predictions:** Since the model shows lower precision and recall for class 1, further research and refinement are needed. Users could focus on gathering more data on customers who churn or applying different modeling techniques to improve these metrics.

**Incorporating Real-Time Data:** Extending the project to include real-time data analysis could enhance the model's predictive capabilities. This would allow businesses to intervene more quickly when a customer shows signs of potential churn.

**Integration with CRM Systems:** Integrating this model with existing Customer Relationship Management (CRM) systems can automate the churn prediction process, making it a seamless part of the customer management workflow.

**Exploring Additional Variables:** Investigating other variables or features that might influence churn, such as customer satisfaction scores, service usage patterns, or competitive offers, could enhance the model's accuracy and provide deeper insights.

**Predictive Maintenance:** In industries like telecommunications or software services, the model could be used to predict not just customer churn but also areas where service disruptions might lead to increased churn, allowing for predictive maintenance.

**Developing Customized Retention Offers:** Based on the model's insights, companies can create customized retention offers or loyalty programs specifically designed to engage customers who are predicted to churn.

**Longitudinal Analysis:** Conducting longitudinal studies to track the effectiveness of different retention strategies over time, as suggested by the model, could provide valuable feedback to continually refine the model and the strategies based on it.

## Conclusion

The optimized SVC model shows promising results in predicting customer churn. However, the noted discrepancy in performance between classes necessitates further refinements. Continuous evaluation and adaptation of the model are crucial for maintaining its relevance and accuracy in real-world scenarios.

The Streamlit web application serves as an accessible and efficient tool for both individual and batch customer churn predictions, leveraging the power of the optimized SVC model. This application is designed to cater to diverse user needs, offering simplicity in single predictions and efficiency in handling bulk data, accompanied by insightful visualizations. It stands as a valuable asset for businesses looking to harness machine learning for customer retention strategies.