# Review of: "Bank Customer Churn Prediction Using SMOTE: A Comparative Analysis"

Divya S[1]

1 University of Buckingham

**Data Specifics and Privacy:**

Could you provide more details about the specific attributes included in the Kaggle Bank Customer Churn Dataset used for this study? How were customer privacy and data protection concerns addressed while conducting this analysis?

**Algorithmic Details:**

For the Genetic Algorithm used in feature selection, what was the basis for choosing the specific parameters (e.g., population size, number of generations)? How did these parameters influence the feature selection process and, subsequently, the performance of the classification models?

**Comparison of Classification Algorithms:**

The KNN model showed superior performance in your study. Could you elaborate on the hyperparameter tuning process for the KNN and other classification models? How sensitive were the models' performances to changes in these hyperparameters?

**Impact of SMOTE:**

Beyond the improved balance in the dataset, how did the application of SMOTE affect the overall quality of the dataset? Were there any observed drawbacks or potential biases introduced by using synthetic samples?

**Feature Importance Analysis:**

In the context of the Genetic Algorithm's application for feature selection, which features were identified as most predictive of customer churn, and why? How did the importance of these features compare across the different models used in the study?

**Robustness and Validation:**

How was the robustness of the proposed models tested against real-world scenarios, particularly in terms of model performance over time and across different customer segments?

**Data Specifics and Privacy:**

Could you provide more details about the specific attributes included in the Kaggle Bank Customer Churn Dataset used for this study? How were customer privacy and data protection concerns addressed while conducting this analysis?

**Algorithmic Details:**

For the Genetic Algorithm used in feature selection, what was the basis for choosing the specific parameters (e.g., population size, number of generations)? How did these parameters influence the feature selection process and, subsequently, the performance of the classification models?

**Comparison of Classification Algorithms:**

The KNN model showed superior performance in your study. Could you elaborate on the hyperparameter tuning process for the KNN and other classification models? How sensitive were the models' performances to changes in these hyperparameters?

**Impact of SMOTE:**

Beyond the improved balance in the dataset, how did the application of SMOTE affect the overall quality of the dataset? Were there any observed drawbacks or potential biases introduced by using synthetic samples?

**Feature Importance Analysis:**

In the context of the Genetic Algorithm's application for feature selection, which features were identified as most predictive of customer churn, and why? How did the importance of these features compare across the different models used in the study?

**Robustness and Validation:**

How was the robustness of the proposed models tested against real-world scenarios, particularly in terms of model performance over time and across different customer segments or geographic locations?

**Practical Implementation Considerations:**

What challenges do you anticipate in implementing the proposed churn prediction models in actual banking operations? How can these models be integrated with existing Customer Relationship Management (CRM) systems?

**Ethical Considerations:**

Were any ethical considerations taken into account while developing and testing these churn prediction models, especially regarding the potential impact on customers identified as likely to churn?

**Future Research Directions:**

You mentioned examining customer churn characteristics in future research. Are there specific aspects or customer behaviors you aim to explore? How do you plan to incorporate these findings into refining churn prediction models?s or

geographic locations?

**Practical Implementation Considerations:**

What challenges do you anticipate in implementing the proposed churn prediction models in actual banking operations? How can these models be integrated with existing Customer Relationship Management (CRM) systems?

**Ethical Considerations:**

Were any ethical considerations taken into account while developing and testing these churn prediction models, especially regarding the potential impact on customers identified as likely to churn?

**Future Research Directions:**

You mentioned examining customer churn characteristics in future research. Are there specific aspects or customer behaviors you aim to explore? How do you plan to incorporate these findings into refining churn prediction models?