PARUL UNIVERSITY

FACULTY OF ENGINEERING AND TECHNOLOGY

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

CRITICAL EVALUATION OF JOURNAL PAPER

This activity is an individual activity. Each student has to refer FIVE research papers related to their project title and each paper is to be reviewed as per below. Student need to prepare word file for each paper evaluation and submit it.

1. Student Details

Student Name:	**Sane Harsh Balkrishna**		
Enrollment No	**2203031250084**		
Title of Journal Paper	**Handling Class Imbalance in Customer Churn Prediction**	**Branch:**	**CSE-BDA**
Authors	J. Burez, D. Van den Poel		
Journal / Conference:	Expert Systems with Applications		
Volume / Issue	36	**Pages:**	11

2. Dissection of Paper

Section 01: Abstract / Introduction (Read the abstract and answer the following questions)

1 What is the objective of the Paper?

Ans: 1. The paper aims to address the issue of class imbalance in customer churn prediction models.

2. It evaluates different techniques such as under-sampling, over-sampling, and cost-sensitive learning to improve predictive performance.

2 What are the main results mentioned in the abstract?

Ans: 1. Under-sampling techniques improved churn prediction accuracy when evaluated using AUC.

2. Weighted Random Forests performed better than standard classifiers.

3. Boosting methods proved to be robust but did not always outperform other techniques.

3 What rational is given by the authors, attributing importance to the research problem?

Ans: 1. Customer churn is a rare event, making it difficult for standard machine learning models to predict accurately.

2. Class imbalance can lead to biased models that favor the majority class, reducing the ability to detect churners.

3. Addressing class imbalance ensures more reliable churn predictions, helping businesses make data-driven decisions.

4 How many earlier works are cited by the authors, and what are the perceived drawbacks of these earlier works?

Ans: 1. Several studies are cited, but most do not adequately handle class imbalance.

2. Previous models tend to use standard accuracy as an evaluation metric, which is not suitable for imbalanced datasets.

Section 02: Methodology

1 Describe the methodology is used by author(s) to address the research problem?

Ans: 1. The study compares various sampling techniques, including random under-sampling, over-sampling, and advanced sampling methods like CUBE.

2. Cost-sensitive learning approaches, including Weighted Random Forests, were also tested.

3. AUC (Area Under Curve) was used as the primary evaluation metric instead of standard accuracy.

2 In what way the methodology used by the authors is relevant to the methodology you proposed to adopt?

Ans: 1. The study provides insights into handling class imbalance, which is crucial for improving churn prediction accuracy.
2. The use of AUC as an evaluation metric ensures a more reliable assessment of model performance in imbalanced datasets.

Section 03: Results and conclusions

1 What are the variables used for the analysis

Ans: Customer demographics, contract details, service usage, and historical churn behavior.

2 List the results obtained by the authors.

Ans: 1. Under-sampling methods improved predictive performance compared to using imbalanced data.

2. Weighted Random Forests performed better than standard Random Forests and Logistic Regression.

3. Boosting algorithms were robust but did not always outperform cost-sensitive methods.

3 What are the conclusions drawn by the authors from the study.

Ans: 1. Addressing class imbalance is crucial for accurate churn prediction.

2. Cost-sensitive learning and sampling techniques can significantly improve model performance.

3. Businesses should prioritize models that optimize for recall and precision instead of overall accuracy.

Write a critical analysis of the paper (about 200 words)

This paper presents a novel approach to churn prediction by integrating privacy-preserving feature engineering with edge computing. The proposed CANCEL method extracts user behavior attributes locally, ensuring data privacy while maintaining high predictive accuracy. The study effectively highlights the growing concern of data privacy in AI-driven customer retention strategies.

One of the significant strengths of the paper is its focus on privacy, which is often overlooked in traditional churn prediction models. The evaluation demonstrates that CANCEL performs competitively compared to centralized models, making it a viable solution for industries handling sensitive user data. However, a limitation of this study is the computational constraints associated with edge devices, which may impact model scalability.

Future research could explore hybrid approaches that combine edge computing with federated learning to further enhance efficiency and scalability. Overall, the study provides valuable insights into privacy-aware machine learning applications, making it relevant for organizations prioritizing data security in churn prediction models.