Assignment 2  
Classification Models

short line

Ngoc Quang Vinh Pham

Student ID: 25100660  
April 04th, 2024

36106 - Machine Learning Algorithms and Applications

Master of Data Science and Innovation

University of Technology of Sydney

Table of Contents

[**1. Executive Summary 2**](#_xvtyf97iwmd0)

[**2. Business Understanding 3**](#_if9iwijz2pg3)

[a. Business Use Cases 3](#_bnjfuh5ewogw)

[**3. Data Understanding 4**](#_yg03y3w5hig9)

[**4. Data Preparation 5**](#_tj77rrmixzhi)

[**5. Modeling 6**](#_6vwh14nc44ga)

[a. Approach 1 6](#_uv8sn0ik6017)

[b. Approach 2 6](#_wtwxuxemqtvc)

[c. Approach 3 6](#_nrl8g3ri789x)

[**6. Evaluation 8**](#_mngiyx5tet6w)

[a. Evaluation Metrics 8](#_qx20m6hrk8uo)

[b. Results and Analysis 8](#_25v57zsj5m7g)

[c. Business Impact and Benefits 8](#_hkob76wu4d6q)

[d. Data Privacy and Ethical Concerns 9](#_uf1z6gbsejg6)

[**7. Deployment 10**](#_8rqux13i2o01)

[**8. Conclusion 11**](#_9hr11g79asdx)

[**9. References 12**](#_ma3y8ytjvp1y)

# Business Understanding

## Business Use Cases

* Describe the specific business use cases or scenarios where the project is applied.
* Discuss the challenges or opportunities that motivated the project.

Instructions: Describe the business use cases or scenarios where the project is applied. Discuss the challenges or opportunities that motivated the project, explaining why machine learning algorithms are relevant in this context.

1. Key Objectives

* Specify the key objectives or goals of the project.
* Identify the stakeholders and their requirements.
* Explain how the project aims to address these requirements.

Instructions: Specify the key objectives or goals of the project, highlighting the desired outcomes. Identify the stakeholders involved and their specific requirements. Explain how the project aims to address these requirements through the use of machine learning algorithms.

short dash

# Data Understanding

* Provide insights into the dataset used for the project.
* Describe the data sources, data collection methods, and any data limitations.
* Discuss the variables/features present in the dataset and their significance.

Instructions: Describe the dataset used for the project, including its sources and any limitations. Discuss the variables or features present in the dataset and their relevance to the project. Include any exploratory data analysis conducted to understand the data better.

short dash

# Data Preparation

* Describe the steps taken to prepare the data for modeling.
* Discuss the data cleaning, preprocessing, and feature engineering techniques applied.
* Document any handling of missing values, outliers, or imbalanced data.

Instructions: Describe the data preparation steps taken before modeling. Include details about data cleaning, preprocessing, and feature engineering techniques applied. Explain how missing values, outliers, or imbalanced data were handled and any transformations performed on the dataset.

short dash

# Modeling

* Describe the machine learning algorithms used for modeling.
* Discuss the rationale behind selecting these algorithms.
* Explain the parameter tuning and model selection process.

Instructions: Describe the machine learning algorithms used for modeling, providing a rationale for their selection based on the project goals. Explain the process of parameter tuning and model selection. Include details about the algorithms' implementation and any considerations made during the modeling phase.

short dash

# Evaluation

## Results and Analysis

* Present the results of the model evaluation, including accuracy, precision, recall, F1-score, etc.
* Analyze and compare the performance of each model.
* Discuss the key insights gained during the experimentation phases.

Instructions: Present the results of the model evaluation, including accuracy, precision, recall, F1-score, or any other relevant metrics. Analyze and compare the performance of each model, highlighting the key insights gained during the experimentation phases. Discuss the implications of these insights on the project's goals and potential areas for further improvement.

## Business Impact and Benefits

* Assess the impact and benefits of the final model on the business use cases.
* Discuss how the model contributes to solving the identified challenges or exploiting opportunities.
* Quantify the improvements achieved and the potential value generated.

Instructions: Assess and discuss the impact and benefits of the final model on the identified business use cases. Explain how the model contributes to solving the identified challenges or exploiting opportunities. Quantify the improvements achieved and discuss the potential value generated by the model.

## Data Privacy and Ethical Concerns

* Assess the data privacy implications of the project.
* Discuss any ethical concerns related to data collection, usage, or model deployment.
* Address steps taken to ensure data privacy and ethical considerations.
* Assess potential negative impacts and risks for Indigenous people

Instructions: Assess the data privacy implications of the project, considering any sensitive information or privacy concerns related to data collection, usage, or model deployment. Discuss any ethical concerns and considerations. Address the steps taken to ensure data privacy and mitigate ethical concerns.

short dash

# Conclusion

* Summarize the key findings, insights, and outcomes of the project.
* Reflect on the project's success in achieving its goals and meeting stakeholders' requirements.
* Discuss any future work, recommendations, or next steps based on the project's outcomes.

Instructions: Summarize the key findings, insights, and outcomes of the project. Reflect on the project's success in achieving its goals and meeting stakeholders' requirements. Discuss any future work, recommendations, or next steps based on the project's outcomes.

short dash

# References

* Include a list of references used throughout the project report.

Instructions: Include a list of references used throughout the project report, following the appropriate citation style.

<https://pranav-c.medium.com/smote-vs-smote-enn-which-is-more-effective-for-churn-prediction-in-imbalanced-banking-data-b289414366a0>

<https://mriquestions.com/frequency-encoding.html#:~:text=Frequency%2Dencoding%20of%20spatial%20position,as%20a%20function%20of%20position>.

<https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222>

<https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/>

<https://realpython.com/python-counter/>

<https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/>

<https://scikit-learn.org/stable/modules/cross_validation.html>

<https://machinelearningcoban.com/2017/08/31/evaluation/>

short dash