Bank Marketing Analysis

Ashel Munemo (am2704)

Introduction

- Objective:
  - Predict whether a client will subscribe to a term deposit.
  - Identify key predictors which influence subscription likelihood.
- Data source:
  - Bank marketing dataset created by Paulo Cortez (Univ. Minho) and Sérgio Moro (ISCTE-IUL) in 2012
- Overview:
  - The data is related with direct marketing campaigns of a Portuguese banking institution.
  - The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank deposit) would be subscribed.

EDA Insights

- The aim was to identify key patterns and preliminary relationships that may influence a client’s decision to subscribe.
- Across all variable categories the number of non-subscribers was consistently higher than subscribers.
- T-test were performed to compare the mean of numerical variables, and the results were significant suggesting that these variables are likely important for predicting client subscriptions.
- Chi-squared test were also conducted on categorical variable to examine the association between each categorical variable and the subscription target, all variables were significant indicating that they are likely important for prediction.

Models & Performance

- The classification models used to predict whether a client subscribes to a term deposit include:
  - Logistic regression, Decision tree, Boosting, and Deep learning
- To evaluate and compare the predictive performance of the models, key metrics are shown below

Model	**AUC**	**Sensitivity (recall)**	**Specificity**	**Precision**	**Accuracy**	**Balanced Accuracy**
Logistic regression	0.9027	0.681	0.910	0.496	0.883	0.795
Decision tree	0.7827	0.944	0.475	0.932	0.890	0.709
Boosting	0.9295	0.490	0.964	0.641	0.909	0.727
Deep learning	0.9086	0.776	0.879	0.455	0.867	0.827

Results

- The comparison on the classification models shows distinct trade-offs between their performance metrics
- Deep learning offers the best trade-off between detecting subscribers and non-subscribers.
- Boosting is best for non-subscribers, and the decision tree is best at identifying subscribers but at a cost of misclassifying many non-subscribers.
- Logistic regression has moderate performance, showing it can detect subscribers and non-subscribers equitably while maintaining good accuracy in identifying non-subscribers.

Conclusion

- In conclusion, the analysis revealed several key insights for predicting client subscriptions:
  - certain client characteristics strongly influence subscription likelihood, for example, clients with successful previous campaigns are more likely to subscribe to a term deposit.
- Deep learning emerged as the most effective mode, providing the best balance between sensitivity and specificity, along with strong overall AUC and the highest balanced accuracy.
- Limitation:
  - The relatively imbalanced dataset may affect the possibility of improving model performance further