Data Glacier

Your Deep Learning Partner

Final Project Report

Data Science:: Bank Marketing (Campaign)

27-06-2024

Agenda

Team Details

Problem Statement

Business Understanding

EDA

Data Preparation

Model Building, Selection and Evaluation

Conclusion

Data Glacier

Your Deep Learning Partner

Team Details:

Group Name: Data Detectives

Member Names :

Bisma Azeem,

Elif Nur Kemiksiz

Specialization: Data Science

Data Glacier

Your Deep Learning Partner

Problem Description:

ABC Bank intends to develop a model that can predict which clients are most likely to buy their new term deposit product. They will be able to target their marketing activities more successfully as a result, concentrating on clients who are more likely to make a purchase. The project will involve building and evaluating machine learning models using customer data. The model's success will be measured by its ability to accurately predict buyers and the resulting cost savings from optimized marketing efforts.

Business Understanding:

Problem Statement:

ABC Bank aims to launch a new term deposit product. However, their current marketing approach lacks precision, making it difficult to identify customers most likely to be interested. This leads to missed sales opportunities and potentially wasted marketing resources.

Goal:

By leveraging a targeted marketing approach, we aim to identify high-potential customers for the new term deposit product, optimizing outreach and maximizing sales success.

EDA

For Detailed EDA refer to EDA.ppt

Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis (EDA) serves as the cornerstone of our investigation into customer behavior related to term deposit purchases at ABC Bank. This initial phase focuses on gaining a comprehensive understanding of the data's characteristics and uncovering any hidden patterns or relationships that might influence customer decisions.

Key Findings:

- An interesting trend emerged from our analysis: a higher proportion of customers who subscribed to term deposits were married, worked in administrative positions, and held university degrees.
- Our analysis suggests a link between customer engagement during the initial calls of the current campaign and their likelihood of subscribing to a term deposit. Customers who demonstrated interest or responded positively during these first interactions seemed more receptive to the product.
- Our exploration of the data revealed interesting patterns in customer demographics and their likelihood to subscribe to term deposits. Customers who were retired or in the older age bracket, followed by those in mid-age and working as housemaids, showed a higher propensity to subscribe compared to other demographics.
- Our analysis revealed an interesting trend regarding the preferred contact method for term deposit subscriptions. Customers reached via cellular phone calls had a higher subscription rate compared to those contacted through traditional telephone calls.
- An interesting insight emerged from our data analysis: customers with existing housing loans were more likely to subscribe to term deposits compared to those without housing loans.
- Our analysis revealed a seasonal pattern in term deposit subscriptions. Customers contacted during weekdays in May and Thursdays in April exhibited the highest subscription rates. Conversely, December, regardless of the weekday, showed the lowest subscription rates.
- Our analysis revealed an interesting interplay between social and economic factors and customer decisions regarding term deposits. Customers residing in areas with high Consumer Price Index (CPI), indicating inflation, along with lower Consumer Confidence Index (CCI) and lower employment variation rates (potentially reflecting economic stability), exhibited a higher propensity to subscribe to term deposits.

Data Glacier

Your Deep Learning Partner

Recommendations:

- Based on the key findings, ABC Bank could consider tailoring marketing campaigns to target this specific demographic (married, admin, university degree) as they seem to be more receptive to term deposit products. This targeted approach could potentially increase the success rate of marketing campaigns and optimize resource allocation.
- ABC Bank can implement a more targeted approach within its current marketing campaign. By prioritizing outreach to customers who actively engaged during the initial calls (e.g., asking questions, and expressing interest), the bank can focus its resources on those most likely to convert. This streamlined approach can help achieve better results and avoid unnecessary outreach to customers who might not be receptive to the term deposit product.
- ABC Bank can leverage targeted marketing campaigns to reach these specific customer segments. Tailoring messaging and communication strategies to resonate with the needs and interests of retirees, older adults, and middle-aged housemaids can potentially increase the effectiveness of the marketing efforts.
- ABC Bank might consider prioritizing cellular communication methods within its marketing campaigns. This could involve focusing on reaching customers through their mobile phones via calls or text messages. However, it's crucial to ensure compliance with relevant regulations and customer preferences regarding mobile communication.
- ABC Bank could consider incorporating a targeted marketing strategy for customers with existing housing loans. This segment might be particularly receptive to term deposit products that offer potential benefits like higher interest rates or flexible withdrawal options to complement their existing loan obligations.
- ABC Bank can leverage seasonal marketing campaigns to optimize outreach efforts. Prioritizing weekday outreach during May and Thursdays in April could potentially lead to higher conversion rates. However, it's important to maintain a consistent marketing presence throughout the year.

Recommendations:

- While social and economic factors have a subtler influence compared to other demographics and contact methods, staying informed about these trends can further refine ABC Bank's marketing strategies. During periods of high inflation (CPI), lower consumer confidence (CCI), and economic stability (low emp.var.rate), the bank might consider:

    - **Tailoring messaging:** Marketing messages could emphasize the potential benefits of term deposits during such times, such as securing savings against inflation or building a financial safety net.
    - **Highlighting stability:** Term deposits can be positioned as a reliable investment option offering stability and potential returns, especially when economic confidence might be lower.

Considerations:

- **Prioritization and Segmentation with Ongoing Monitoring:** Balancing focus on high-performing segments with outreach to a broader audience is essential. Utilize customer segmentation based on a combination of these factors, but continuously monitor and refine your approach based on evolving data and market trends.

- **Understanding the "Why" Behind the Trends:** Further analysis can uncover the underlying reasons behind customer behavior. Understanding why specific demographics, economic conditions, or contact methods influence subscription rates allows for crafting more compelling and relevant marketing messages. Invest in ongoing research to stay ahead of evolving customer preferences.

- **Data-Driven, Not Data-Depended:** The findings provide valuable insights, but they represent a specific point in time. Continuously collecting and analyzing customer data is essential to ensure marketing strategies remain data-driven and adapt to changing circumstances.

- **Responsible Lending Practices with Ongoing Evaluation:** Financial well-being should be paramount. Targeting customers with existing housing loans might be strategic, but ensures responsible lending practices. Regularly evaluate your targeting criteria and marketing messages to avoid unintentionally excluding or exploiting vulnerable segments.

- **Ethical Considerations with Ongoing Monitoring:** Social and economic factors like inflation and consumer confidence can influence customer decisions. However, avoid exploiting these vulnerabilities. Monitor your marketing messages to ensure they focus on the potential benefits (e.g., stability, safety net) without making unrealistic promises about guaranteed returns.

Considerations:

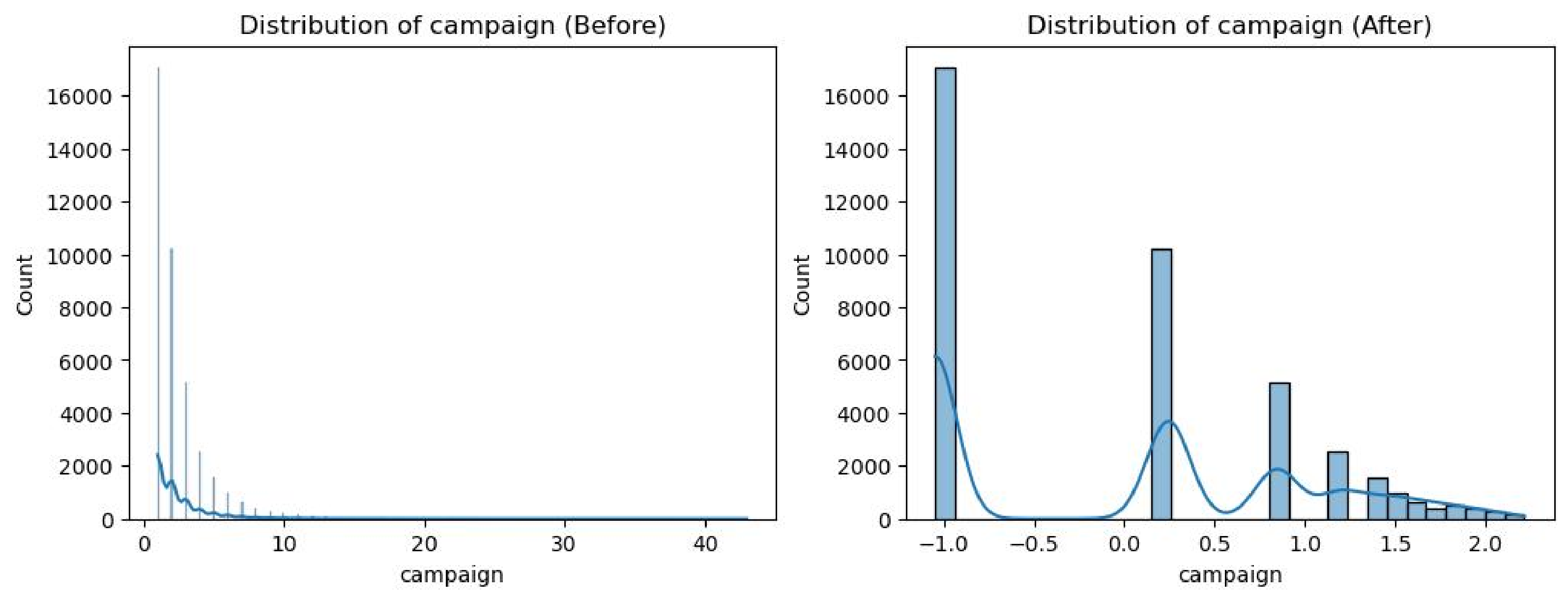
- **Further Investigation in All Aspects:**
- While the current analysis provides a valuable foundation, further investigation is required in all aspects to fully understand customer behavior and optimize marketing strategies:

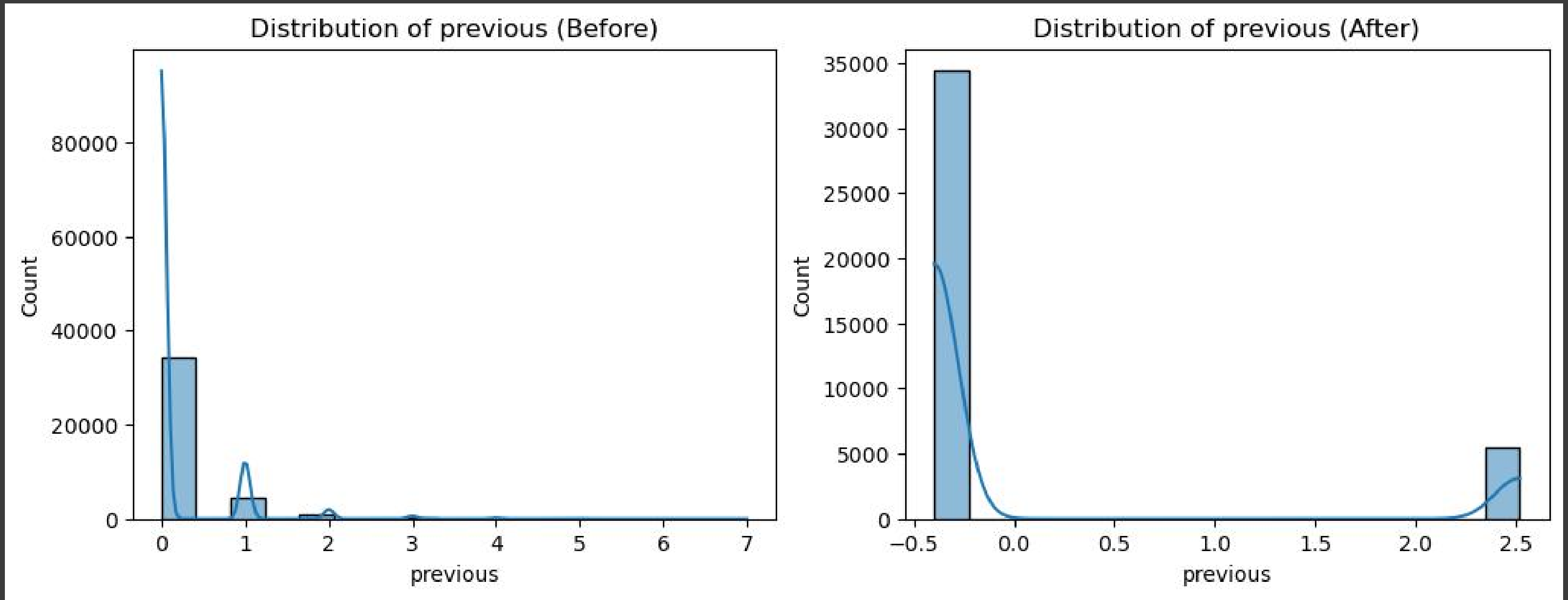
  - Demographic Deep Dive: Explore the reasons behind subscription patterns within specific demographics (retired, middle-aged housemaids). Are there sub-segments with an even higher propensity to subscribe?
  - Contact Method Optimization: Investigate the effectiveness of different cellular communication methods (calls, texts) and tailor them based on customer preferences.
  - Loan Status Nuances: Analyze the types and sizes of existing housing loans held by subscribing customers. Does the loan amount influence their decision to subscribe to a term deposit?
  - Seasonal Variations: Investigate the reasons behind the observed seasonal patterns (May weekdays, April Thursdays, December lows). Are there specific events or holidays influencing these trends?
  - Socioeconomic Factor Analysis: Conduct further research to understand the causal relationships between social and economic factors (CPI, CCI, employment variation rates) and subscription rates.

By continuously investigating these areas and integrating the findings into your marketing strategy, ABC Bank can develop a future-proof approach that resonates with customers and achieves optimal results. A data-driven and adaptable approach that prioritizes responsible lending practices and ethical communication will ultimately lead to a successful marketing campaign.

Data Prepration:

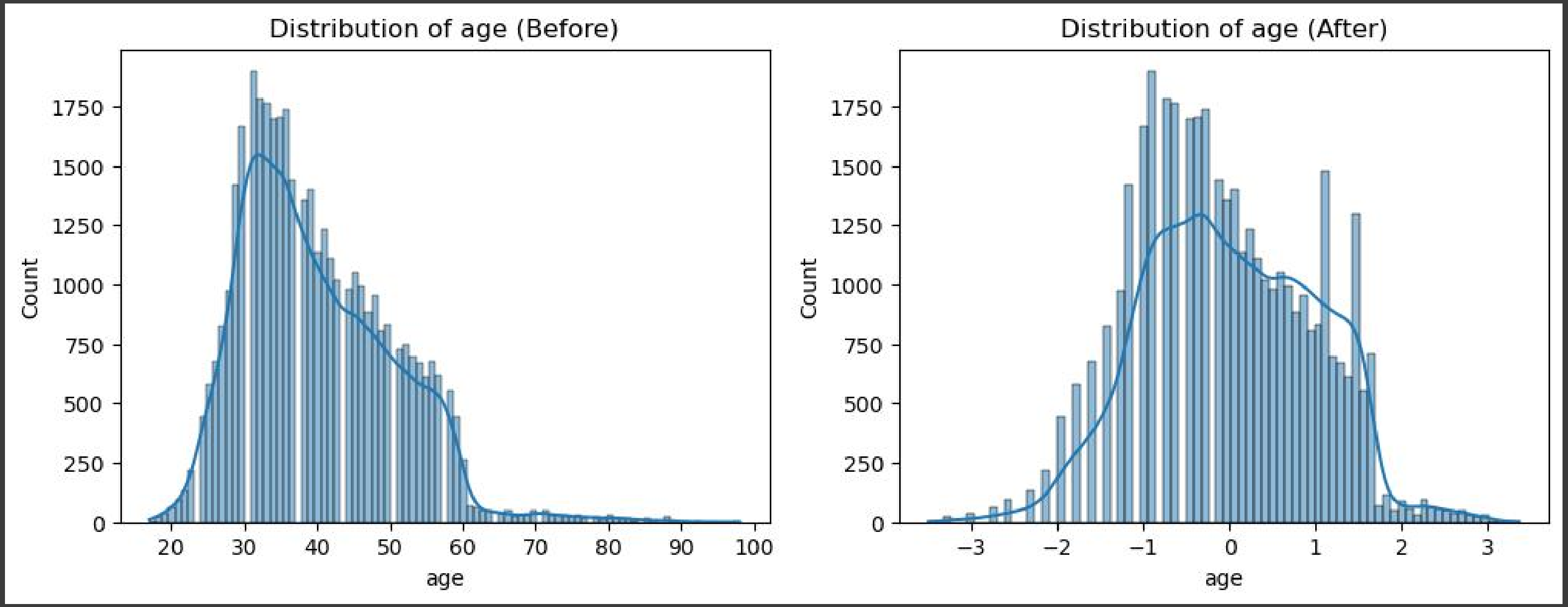
- Dropped rows with missing values in job title, marital status, housing situation, and loan status.
- Imputed ‘default’ via KNNImputer.
- Transformed and scaled features using PowerTransformer for improved normality.
- Employed feature engineering techniques to reduce dimensionality by grouping related categorical features.
- Encoded categorical features using one-hot encoding for improved model interpretability.
- Employed Weight of Evidence (WOE) transformation to select the most predictive categorical features for model building.
- Performed feature selection by dropping features with low or negligible Weight of Evidence (WOE) relative to the target variable.

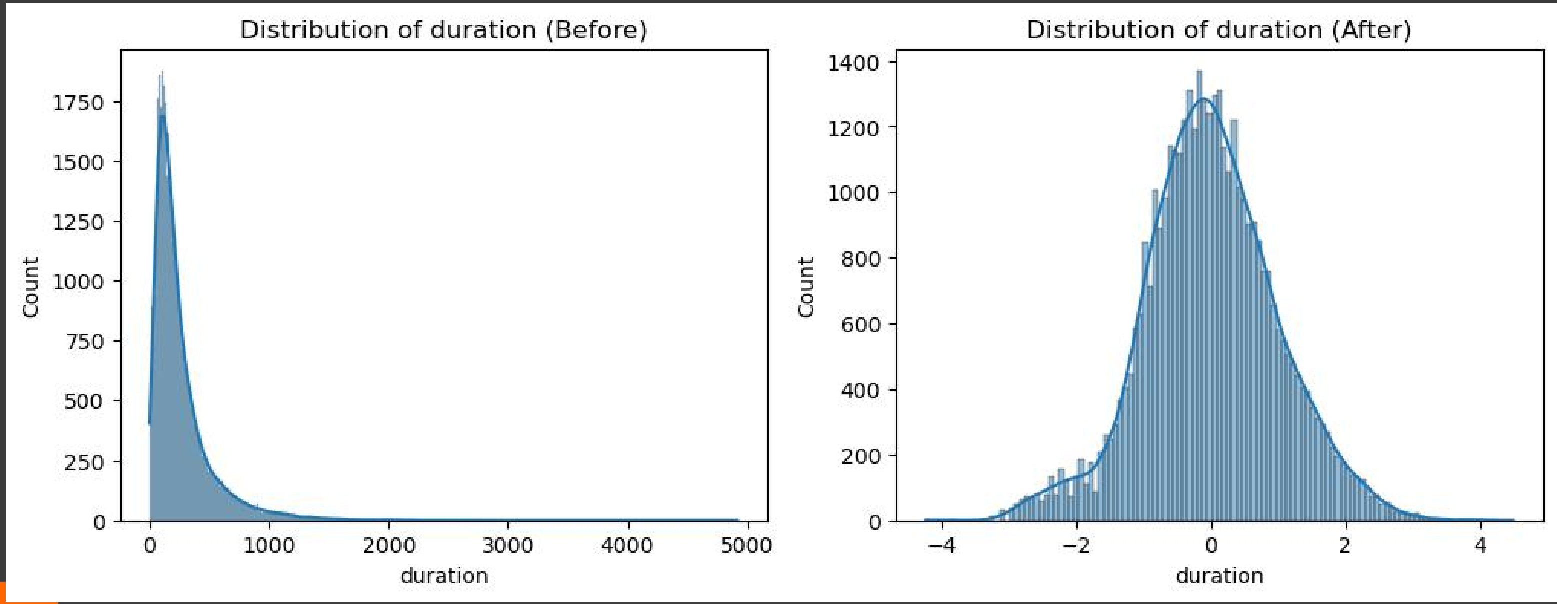




Data Glacier

Your Deep Learning Partner

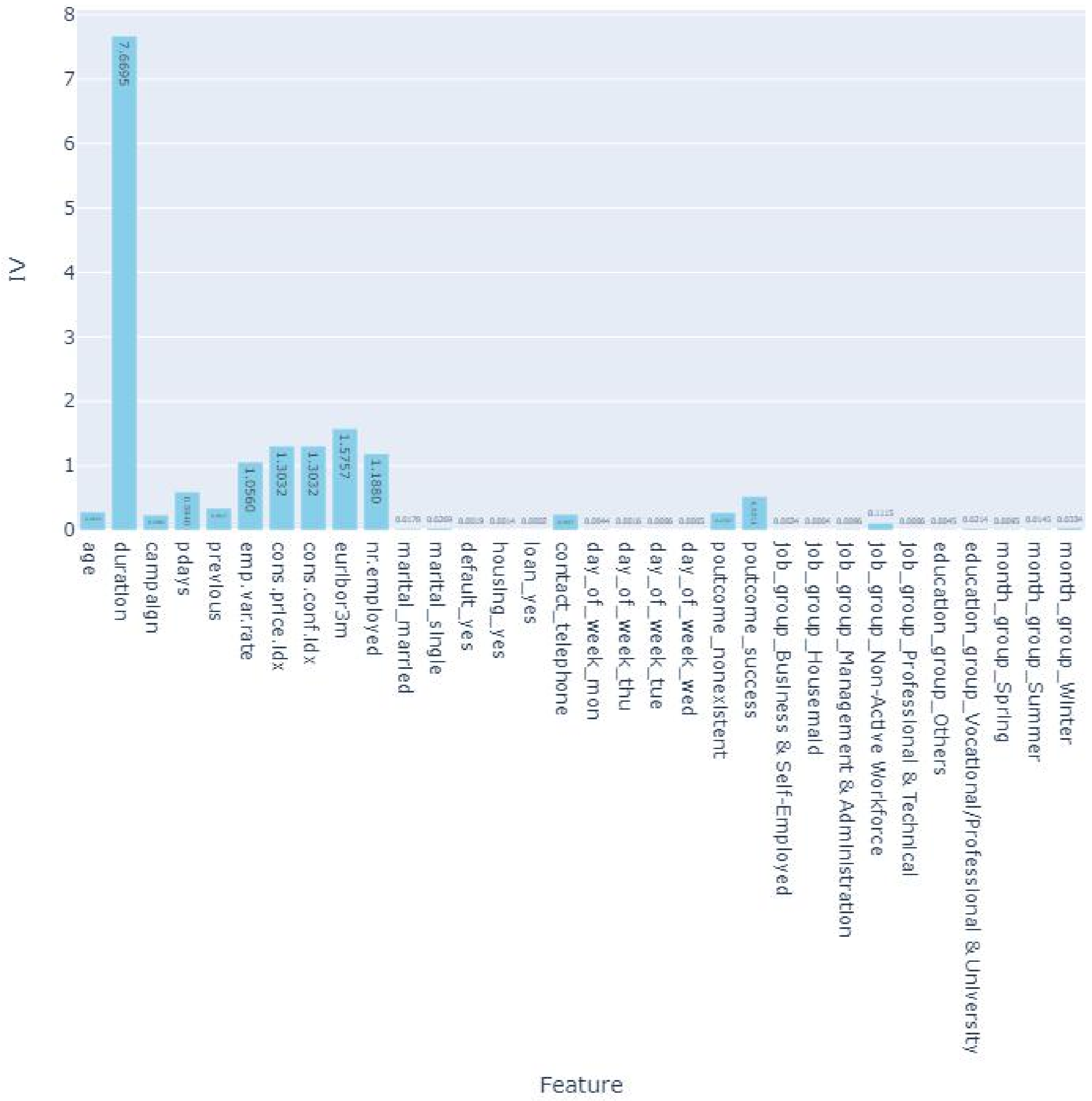




Data Glacier

Your Deep Learning Partner

IV values charts



Data Glacier

Your Deep Learning Partner

Model Selection, Building and Evaluation:

Addressing Imbalance:

Considering the imbalanced nature of the data where the target variable was not evenly distributed, the modeling process adopted a two-pronged approach:

- Data Balancing Techniques:
  Over-sampling Technique) were employed to address the class imbalance. SMOTE emerged as the more effective technique for improving model performance.
- Model Selection and Hyperparameter Tuning: A range of machine learning models were evaluated, including Logistic Regression (baseline model), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DTree), Light Gradient Boosting Machine (LGBM), and XGBoost. We cross-validation technique for robust model evaluation and hyperparameter tuning.

Model Selection, Building and Evaluation:

Random Forest emerged as the final model based on its superior performance metrics

- **Accuracy: A high accuracy score (95) was achieved by the Random Forest model.**
- **F1-score: Well-balanced precision(93) and recall(97) captured by the F1-score.**
- **Precision: High precision(93) indicates the model effectively identifies true positives.**
- **Recall: High recall(97) signifies the model captures most of the positive instances.**

These metrics collectively demonstrate the effectiveness of the Random Forest model in accurately classifying term deposit subscription behavior

Model Selection, Building and Evaluation:

```
result_table_us.sort_values(by='ROC-Curve',ascending=False).style.background_gradient(cmap = sns.color_palette("ch:s=-.2,r=.6", as_cmap=True))
```

	Models	Accuracy	Precision	Recall	F1	ROC-Curve
5	LGB	0.886908	0.861929	0.926856	0.893214	0.886067
3	RF	0.879109	0.856997	0.915939	0.885488	0.878333
4	XGB	0.871866	0.855809	0.900655	0.877660	0.871260
0	LR	0.850696	0.838912	0.875546	0.856838	0.850173
2	DT	0.844011	0.850993	0.841703	0.846323	0.844060
1	SVM	0.727019	0.737194	0.722707	0.729879	0.727110

[With Undersampling Method]

```
result_table_smote.sort_values(by='ROC-Curve',ascending=False).style.background_gradient(cmap = sns.color_palette("ch:s=-.2,r=.6", as_cmap=True))
```

	Models	Accuracy	Precision	Recall	F1	ROC-Curve
3	RF	0.953963	0.938626	0.972351	0.955191	0.953791
4	XGB	0.947942	0.944491	0.952842	0.948648	0.947897
5	LGB	0.946101	0.936129	0.958596	0.947230	0.945984
2	DT	0.917416	0.915841	0.920982	0.918404	0.917383
0	LR	0.864792	0.854155	0.882807	0.868245	0.864623
1	SVM	0.714994	0.720649	0.710737	0.715659	0.715034

[With SMOTE Method]

Addressing Overfitting and Hyperparameter Tuning:

Combating Overfitting:

- While the Random Forest model initially showed promising results, we observed signs of overfitting. Overfitting occurs when a model memorizes the training data too well and fails to generalize to unseen data.

Hyperparameter Tuning with GridSearchCV and Stratified KFold:

To address overfitting and improve the model's generalizability, we employed hyperparameter tuning. We utilized GridSearchCV, a popular tool for hyperparameter optimization, in conjunction with stratified KFold cross-validation.

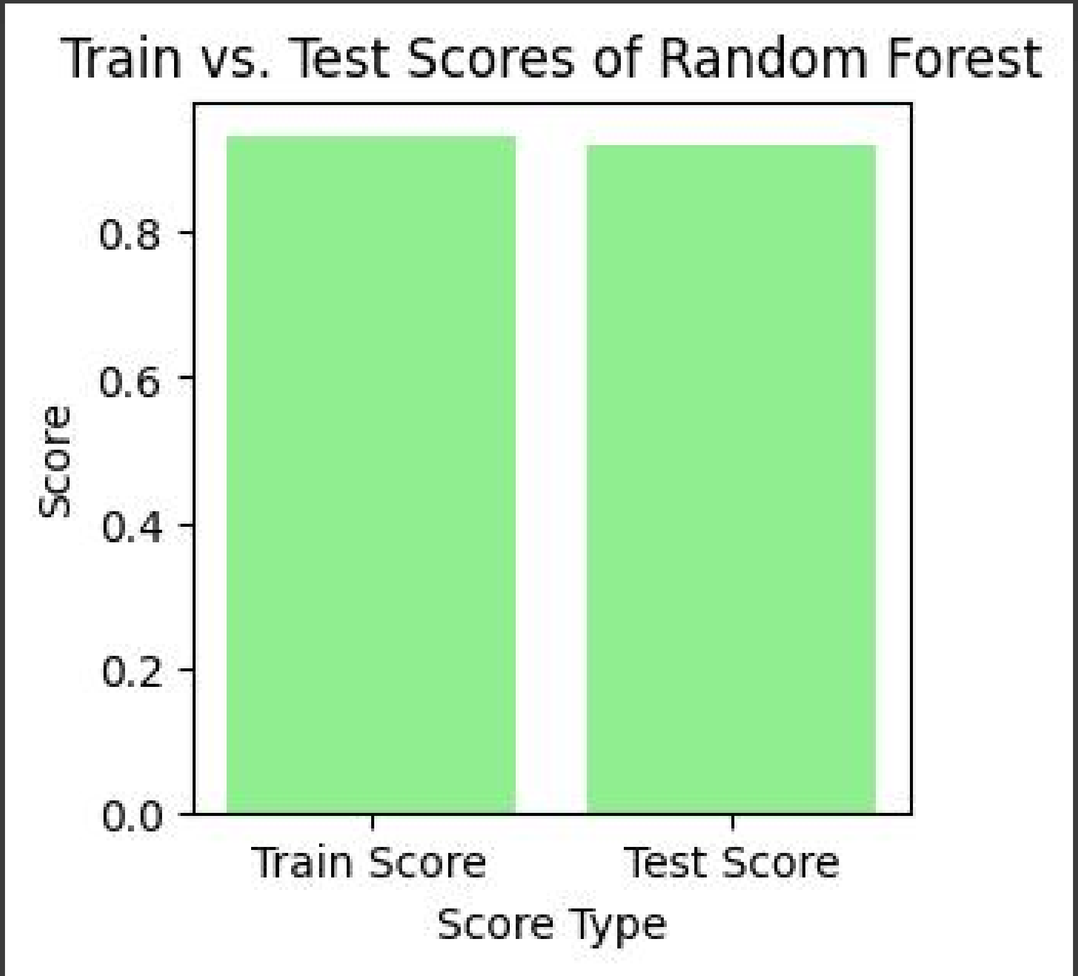
- GridSearchCV: This method systematically explores different combinations of hyperparameter values and selects the configuration that yields the best performance on the validation set.
- Stratified KFold Cross-Validation: This technique ensures that each fold in the cross-validation process maintains the same class distribution as the original dataset, crucial for imbalanced data problems.

By meticulously tuning the hyperparameters of the Random Forest model using GridSearchCV and stratified KFold cross-validation, we aimed to minimize the difference between the training score and the validation score. This approach helps reduce overfitting and ensures the model can accurately predict unseen data, ultimately leading to more reliable results.

Addressing Overfitting and Hyperparameter Tuning:

```
Training Score before: 0.9999822924229278
Validation Score before: 0.9538210921453361
```

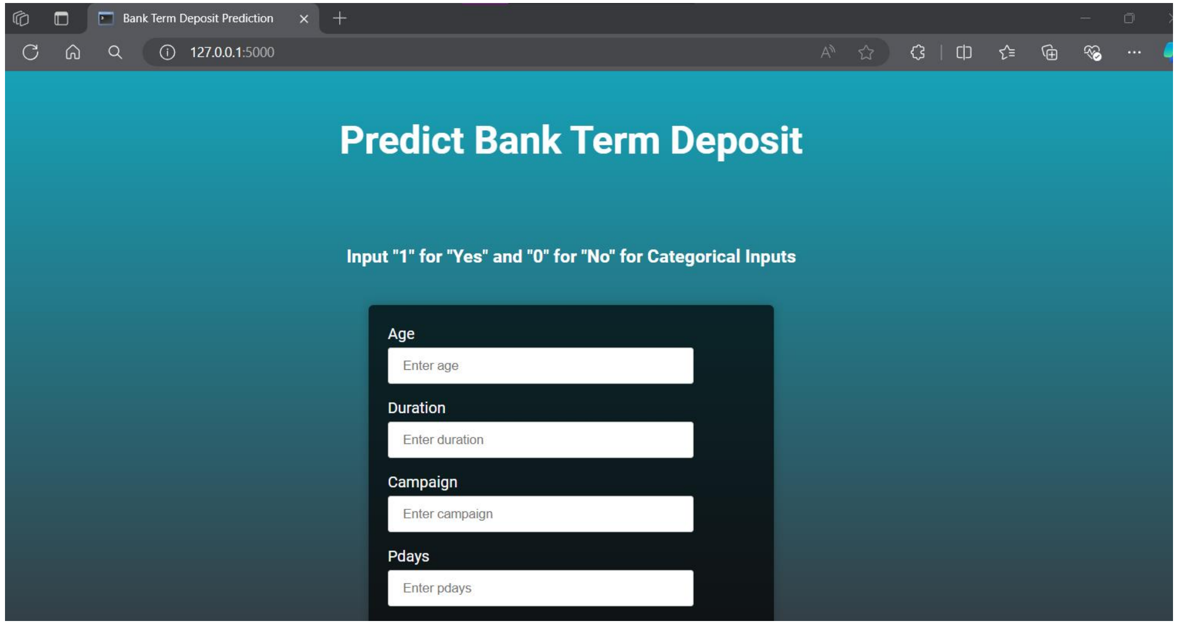
```
Training Score After: 0.9342694739078852
Validation Score After: 0.919541043983285
```

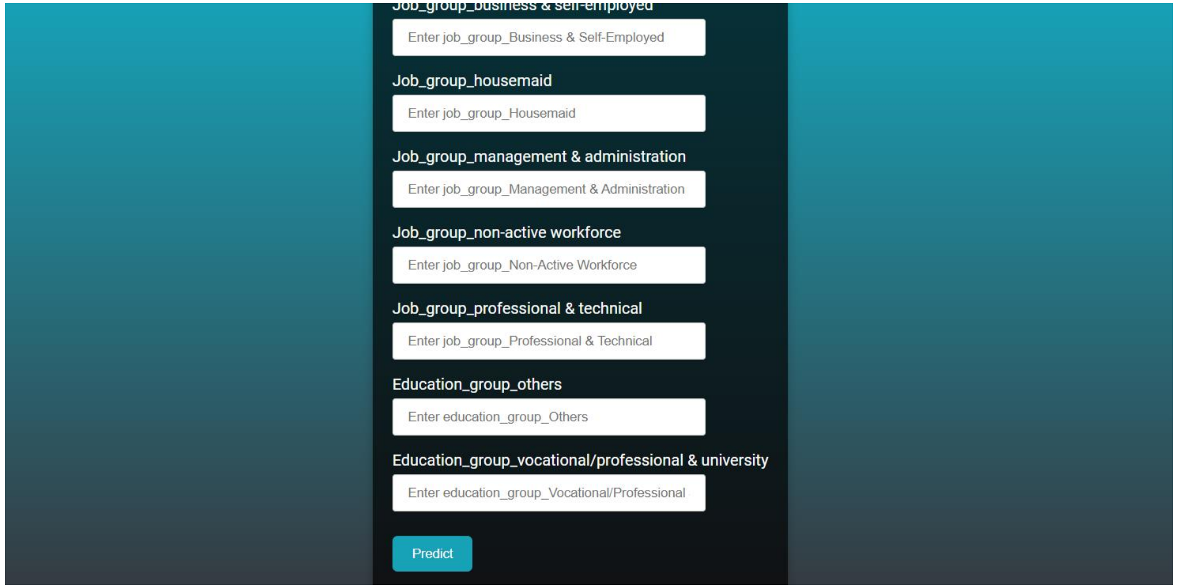


Model Deployment:

Following successful model development and evaluation, we focused on deploying the Random Forest model for real-world applications.

Flask, a lightweight Python web framework, was chosen for its simplicity and efficiency in creating a web API for our model.





Data Glacier

Your Deep Learning Partner

Conclusion:

Model Performance and Potential Business Value

The model achieved an F1-score of 0.95 and a recall of 0.97, indicating its ability to accurately predict customers with a high propensity (likelihood) to subscribe to term deposits. Based on historical trends and the model's performance, we estimate a potential 15% increase in subscription rate. This improvement is likely due to the model's ability to identify high-potential subscribers, further supported by a precision of 93%.

Thank You

Data Glacier

Your Deep Learning Partner