**Business Insights -**

**Variables in the dataset -**

|  |  |  |
| --- | --- | --- |
| S.no | VARIABLE | MEANING |
| 1. | DemoReqPg\_CallClicks\_evt\_count | The event fired for demo calls |
| 2. | air\_purifier\_page\_top | The time spent on the air purifier page |
| 3. | bookdemo\_in\_7days |  |
| 4. | bounces | Total number of single page (or single interaction hit) sessions for the property |
| 5. | checkout\_page\_top | The time spent on the checkout page |
| 6. | contactus\_top | The time spent on the Contact Us page |
| 7.  8. | country | the user’s country type (whether they are domestic or international) |
| 9. | customer\_service\_amc\_login\_top | Time spent on the customer service amc login |
| 10. | customer\_service\_request\_login\_top | Time spent on the customer service request login |
| 11. | demo\_page\_top | Time spent on the demo page |
| 12. | device | The device the user used while working |
| 13. | dsls | Day since the last session |
| 14. | fired\_DemoReqPg\_CallClicks\_evt | Event fired for demo calls |
| 15. | fired\_help\_me\_buy\_evt | Event fired for ‘Help Me Buy’ CTA |
| 16. | fired\_phone\_clicks\_evt | Event fired for ‘Phone Clicks’ CTA |
| 17. | goal3Completions | Book Demo Thank You (Goal 3 Completions) |
| 18. | goal4Completions | Book Demo Thank You (Goal 4 Completions) |
| 19. | goal7Completions | Book Demo Thank You (Goal 7 Completions) |
| 20. | help\_me\_buy\_evt\_count | Event count for ‘Help Me Buy’ CTA |
| 21. | newUser | To check if the user’s new |
| 22. | offer\_page\_top | Time spent on the offers page |
| 23. | pageviews | The number of page views |
| 24. | paid | Whether the user came to the website through a paid campaign |
| 25. | phone\_clicks\_evt\_count | Event count for Phone Clicks CTA |
| 26. | region | User region (derived from their geo location) |
| 27. | security\_solutions\_page\_top | Time spent on the security solutions page |
| 28. | sessionDuration | Total duration (in seconds) of a user’s stay |
| 29. | sessions | Total number of sessions per user |
| 30. | sourceMedium | Combined values of ga: source and ga: medium |
| 31. | storelocator\_top | Time spent on the store locator page |
| 32. | successbookdemo\_top | Time spent on the Success Book Demo page |
| 33. | vacuum\_cleaner\_page\_top | Time spent on the Vacuum Cleaner page |
| 34. | visited\_air\_purifier\_page | If user visited the Air Purifier page |
| 35. | visited\_checkout\_page | If user visited the Checkout page |
| 36. | visited\_contactus | If user visited the Contact Us page |
| 37. | visited\_customer\_service\_amc\_login | If user visited the Customer Service AMC page |
| 38. | visited\_customer\_service\_request\_login | If user visited the Customer Service Request Login page |
| 39. | visited\_demo\_page | If user visited the demo page |
| 40. | visited\_offer\_page | If user visited the offer page |
| 41. | visited\_security\_solutions\_page | If user visited the security solutions page |
| 42. | visited\_storelocator | If user visited the store locator page |
| 43. | visited\_successbookdemo | If user visited the success book demo page |
| 44. | visited\_vacuum\_cleaner\_page | If user visited the Visit Vacuum Cleaner Page |
| 45. | visited\_water\_purifier\_page | If user visited the Water Purifier page |
| 46. | water\_purifier\_page\_top | Time spent on the Water Purifier page |
| 47. | callclick | in\_7days |
| 48. | converted | whether the lead has been converted or not |

**Reference links used for code explanation:**

*Random forest for feature selection:*

[Random Forest Explained. Random Forest explained simply: An easy… | by z\_ai | Towards Data Science](https://towardsdatascience.com/random-forest-explained-7eae084f3ebe)

*Encoding Techniques:*

[What is One Hot Encoding? Why and When Do You Have to Use it? | HackerNoon](https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f)

[When to Use One-Hot Encoding in Deep Learning? (analyticsindiamag.com)](https://analyticsindiamag.com/when-to-use-one-hot-encoding-in-deep-learning/)

[Target Encoding Vs. One-hot Encoding with Simple Examples | by Svideloc | Analytics Vidhya | Medium](https://medium.com/analytics-vidhya/target-encoding-vs-one-hot-encoding-with-simple-examples-276a7e7b3e64)

Data Imbalance

[5 Important Techniques To Process Imbalanced Data In Machine Learning (analyticsindiamag.com)](https://analyticsindiamag.com/5-important-techniques-to-process-imbalanced-data-in-machine-learning/#:~:text=1%20%7C%20Oversampling.%20This%20technique%20is%20used%20to,get%20a%20better%20result%20to%20handle%20imbalance%20data.)

[Handle Class Imbalance using Class Weight - Python - Data Analytics (vitalflux.com)](https://vitalflux.com/class-imbalance-class-weight-python-sklearn/)

[How To Dealing With Imbalanced Classes in Machine Learning (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/)

For this particular case study, we have used the following models:

- Voting

- Logistic Regression

- KNeighbors Classifer

- Decision Tree

- Bagging

- Random Forest

- Boosting

- Adboost

- Gradient Boosting

- XGboost

Which is a better model for sales conversion?

[Convert More Sales Leads With Machine Learning - GormAnalysis](https://www.gormanalysis.com/blog/convert-more-sales-leads-with-machine-learning/)

[Conversion funnel optimization using machine learning (tdcommons.org)](https://www.tdcommons.org/cgi/viewcontent.cgi?article=2128&context=dpubs_series)

[Combine Your Machine Learning Models With Voting | by Aashish Nair | Towards Data Science](https://towardsdatascience.com/combine-your-machine-learning-models-with-voting-fa1b42790d84)

**Codes you can use for studies:**

[GitHub - salinaaaaaa/Machine-Learning-For-Predictive-Lead-Scoring: Predictive Lead Scoring does all the hard work for you by leveraging Machine Learning to provide your sales and marketing team with in-depth customer knowledge and ways to target the hottest and most qualified leads – resulting in saved time and higher revenue streams.](https://github.com/salinaaaaaa/Machine-Learning-For-Predictive-Lead-Scoring)

**Precision Scores for each model as per mean and SD (Where Mean comes first, followed by StD)**

**XGBoost :**

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.08, 0.004) |
| Class weights (scale\_pos\_weight used in absence of class weight hyper parameter) | (0.52, 0.001) |

**Adaboost**:

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.67, 0.012) |
| Without Class weights (no such hyperparameter) | 0.47, 0.34) (std is high as class weight is not applied) |

**Random Forest**:

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.50, 0.0007) |
| Class weights | (0.097, 0.003) |

**Voting:**

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.59, 0.004) |
| Class weights | (0.17, 0.02) |

**KNN:**

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.58, 0.004) |
| Without Class weights (no such hyperparameter) | (0.16, 0.20)(std is high as class weight is not applied) |

**Decision Trees:**

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.64, 0.006) |
| Class weights | (0.097, 0.004) |

**Logistic Regression:**

|  |  |
| --- | --- |
| Under sampling/Over sampling | (0.70, 0.015) |
| Class weights | (0.11, 0.007) |

Finally, we chose **RandomForrest** with selected hyperparameters post randomsearch and gridsearch as our final model as that was giving the most consistent precision on test dataset.

Also, we noticed that Adaboost and XGboost techniques tend to overfit in absence of proper hyperparameter tuning.

**Final Precision\_score** on test dataset = 0.5