ClickWise

CTR Prediction Model

Link to the Notebook: [Notebook](https://colab.research.google.com/drive/1z5PnoVaeh_Ukbsn75i7k1Z-r77Pip93-?usp=sharing)

Link to the GitHub Repo: [Repository](https://github.com/ananya7rai/ClickWise)

## Motivation

In the realm of online advertising, click-through rate (CTR) serves as a crucial metric for assessing the effectiveness of advertisements. Consequently, click prediction systems are vital and extensively utilized for sponsored searches. In this assignment, we are provided with 90 days' worth of Zepto data to construct and evaluate prediction models. Our objective is to develop a CTR prediction model using the provided feature set.

## Dataset

We are provided with a csv file containing 90 days worth of Zepto data with 31 features.

## Proposed architecture

## 

*Fig 1.* Flow Diagram showing the necessary steps required for our project

## Approach

1. **Import libraries and load the dataset :**

* Imported necessary python libraries like Pandas, NumPy, Matplotlib and Seaborn.
* Read the csv file using Pandas.
* Extracted basic information about the dataset like number of columns and rows as well as the data type of different features.
* Displayed statistical information like mean, standard deviation, minimum and maximum value etc. using basic descriptive statistics.

1. **EDA:**

* Checked for missing values and handled them.
* visualized the distribution of the target variable ‘is\_clicked’ to understand the class balance.
* Used Box Plots and Correlation Matrix to understand the relation between features (such as ‘total\_clicks’ and ‘session\_views’) and the target variable.

1. **Feature Engineering:**

* Scaled the features using StandardScaler to ensure they are on a comparable scale.
* Selected important features based on correlation threshold to reduce the dimensionality of the dataset.

1. **Model Building and Evaluation:**

* Trained 3 models : Logistic Regression, Random Forest, and Gradient Boosting.
* Calculated the performance metrics(AUC, accuracy, F1 score) for each model.

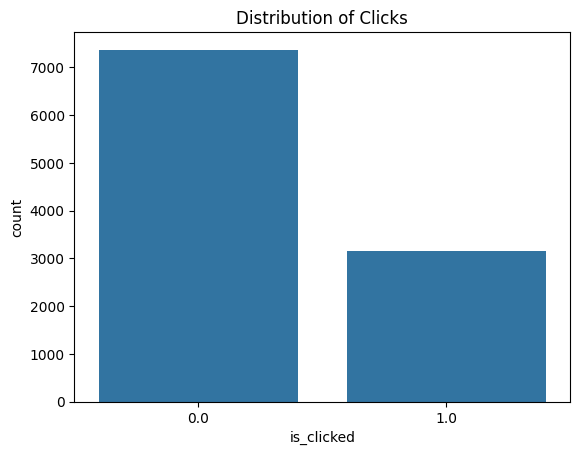
1. **ROC Curve Analysis:**

* Plotted the ROC curve for each model to visualize their performance.
* The Gradient Boosting model shows the highest AUC, indicating better discrimination ability.

1. **Hyperparameter Tuning:**

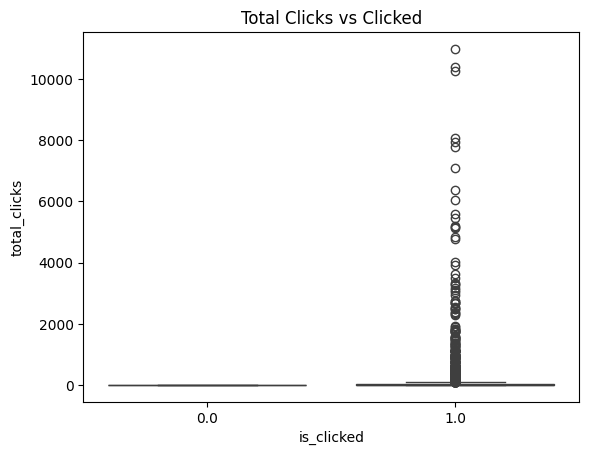
* Performed hyperparameter tuning on the best model (Gradient Boosting) to optimize its performance further.
* The best parameters found through GridSearchCV are used to retrain the model and evaluate it again.

## Results

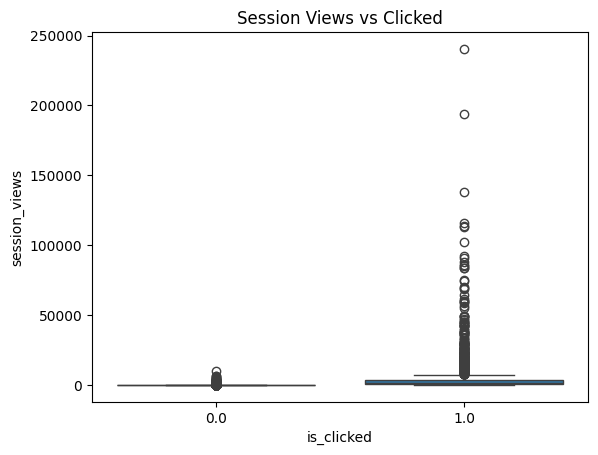


*Fig 2.* Distribution of Clicks

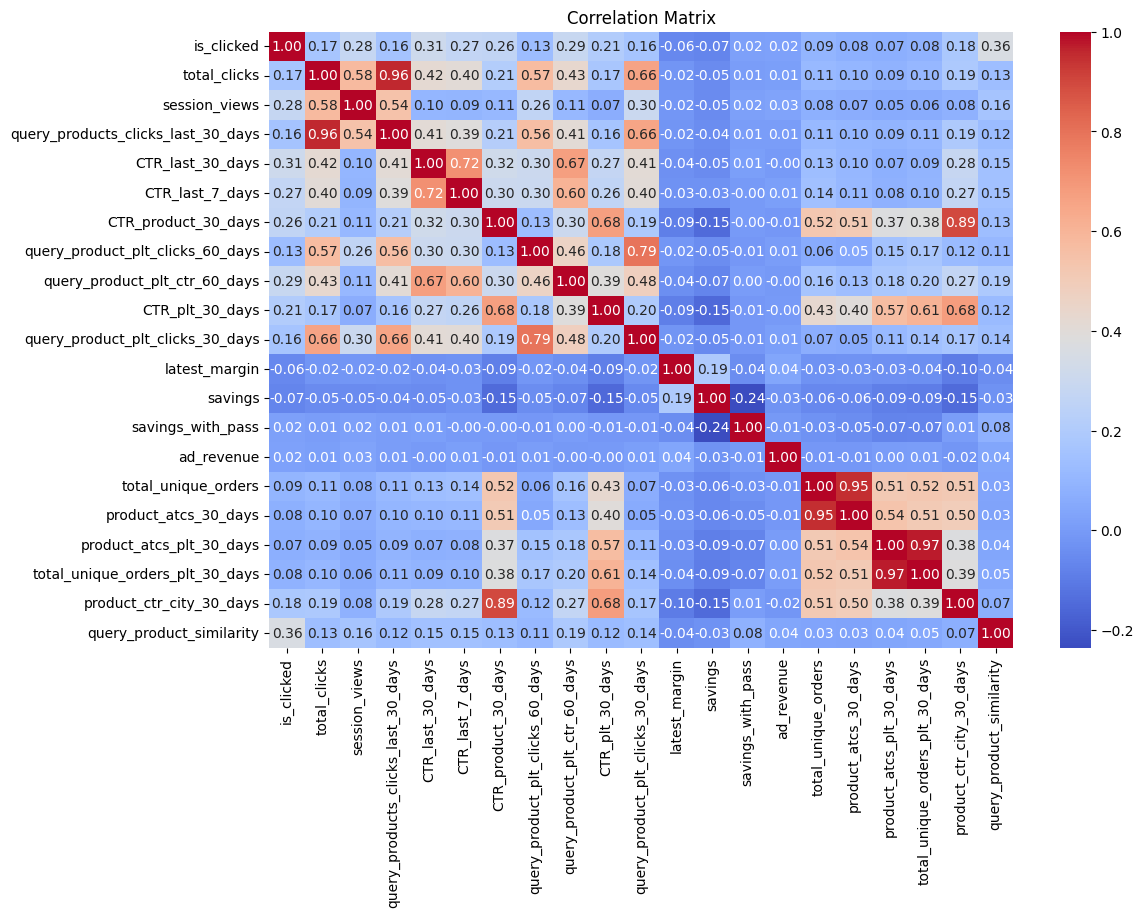
(showing 1.0 as clicked and 0.0 as not clicked)



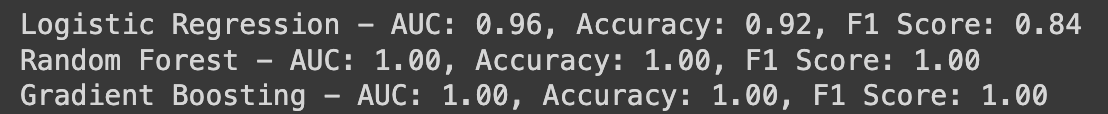
*Fig 3.* Box Plot (total clicks Vs clicks)



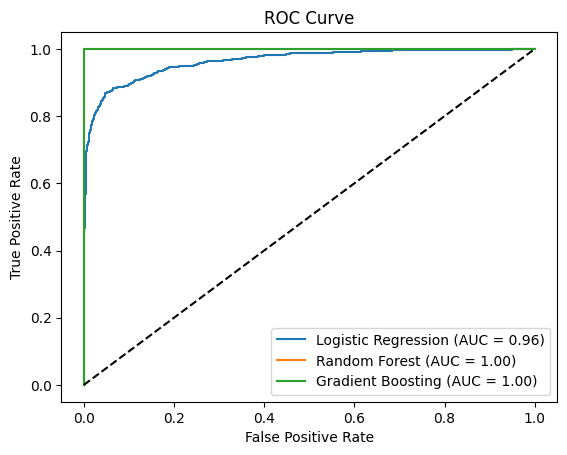
*Fig 4.* Box Plot (session views Vs clicks)



*Fig 5.* Correlation Matrix



*Fig 6.* Evaluation of the models before hyperparameter tuning



*Fig 7.* ROC curve



Fig 8. Evaluation of the best model (Gradient Boosting) after hyperparameter tuning

## 

## Conclusion

| **Metrics** | **Logistic Regression** | **Random Forest** | **Gradient Boosting** |
| --- | --- | --- | --- |
| **AUC** | 0.96 | 1.00 | 1.00 |
| **Accuracy** | 0.92 | 1.00 | 1.00 |
| **F1 score** | 0.86 | 1.00 | 1.00 |

*Table 1.* Evaluation Metrics of different ML models

* Logistic regression is a simple and interpretable model but it showed the lowest performance metrics among the three.
* Random Forest performed better than Logistic Regression, indicating that ensemble methods can capture complex patterns.
* We can conclude that **Gradient Boosting** is the best model to predict the Click through rate using the given data as it achieved the highest AUC, accuracy, and F1 score, demonstrating its effectiveness in handling this problem.
* The Gradient Boosting model, after hyperparameter tuning, achieved an AUC of 1.00, accuracy of 1.00, and F1 score of 1.00.
* These metrics indicate that the Gradient Boosting model provides the best balance between true positive rate and false positive rate, overall accuracy, and the balance between precision and recall.

Thank You!