

CUSTOMER\_RETENTION\_DATASET



NAME OF THE PROJECT

Submitted by:

LODEM RAKESH

YOUR NAME

## **ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

Internship data

# INTRODUCTION

- **Business Problem Framing**

Describe the business problem and how this problem can be related to the real world.

This helps in analysing the business dataset and draw some inference about the ecommerce websites.

- **Conceptual Background of the Domain Problem**

Describe the domain related concepts that you think will be useful for better understanding of the project.

Some knowledge about the e commerce platforms.

- **Review of Literature**

This is a comprehensive summary of the research done on the topic. The review should enumerate, describe, summarize, evaluate and clarify the research done.

Not much research is done.

- **Motivation for the Problem Undertaken**

Describe your objective behind to make this project, this domain and what is the motivation behind.

Check the operating and inference of the ecommerce website.

## **Analytical Problem Framing**

- **Mathematical/ Analytical Modeling of the Problem**

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

Check the null values

Outliers checking

- **Data Sources and their formats**

What are the data sources, their origins, their formats and other details that you find necessary? They can be described here.

Provide a proper data description. You can also add a snapshot of the data.

Internship data..flip robo technologies

- **Data Preprocessing Done**

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

Check the null values

Outliers checking,duplicates checking

- **Data Inputs- Logic- Output Relationships**

Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

We observed some of the columns impacted the output of the data that we got using pca.

- **State the set of assumptions (if any) related to the problem under consideration**

Here, you can describe any presumptions taken by you.

No assumptions taken

- **Hardware and Software Requirements and Tools Used**

Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

Python

Jupyter notebook

## **Model/s Development and Evaluation**

- **Identification of possible problem-solving approaches (methods)**

Describe the approaches you followed, both statistical and analytical, for solving of this problem.

Chi2 test,pca,standardScaler...etc

- **Testing of Identified Approaches (Algorithms)**

Listing down all the algorithms used for the training and testing.

We only used the RandomForestClassifier

- **Run and Evaluate selected models**

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

We selected the best model as RandomForestClassifier and we run the best model.

- **Key Metrics for success in solving problem under consideration**

What were the key metrics used along with justification for using it?  
You may also include statistical metrics used if any.

Accuracy\_score

Cross val score

Confusion matrix

- **Visualizations**

Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

Stripplot

Countplot

Pie chart

Barplot

Violinplot..etc

- **Interpretation of the Results**

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

## **CONCLUSION**

- **Key Findings and Conclusions of the Study**

Describe the key findings, inferences, observations from the whole problem.

Lot of inferences were found during the analysis of the dataset such as most customer preferred e commerce platform ...reasons for the negative reviews for these platforms..why some of the platforms lag behind..

- **Learning Outcomes of the Study in respect of Data Science**

List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

Lot of learning related to the e commerce websites functioning ,their operations..etc

- **Limitations of this work and Scope for Future Work**

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.

We can use all the techniques in future even after data drift.