

High Level Design (HLD)

Google Analytics Customer Revenue Prediction

Revision Number: 2.0

Last date of revision: 26 February 2022

Sugaiel Fathima A

Md Zaid N

Document Version Control

Date Issued	Version	Description	Author
05/02/2023	1.0	Added Introduction and General description	MD Zaid N
08/02/2023	1.1	Added Design details and abstract	Sugaiel Fathima A
11/02/2023	1.2	Created and organized whole	Md Zaid N
14/02/2023	1.3	Updated Design Details	Sugaiel Fathima A
18/02/2023	1.4	Checked and added Performance	Md Zaid N
23/02/2023	1.5	Added Conclusion and References	Sugaiel Fathima A
26/02/2023	1.6	Updated diagrams in Design Details	MD Zaid N

Table of Content

Document Version Control	2
Abstract	4
Introduction	5
1. Why this High-Level Document?	5
2. Scope	5
Definition	5
General Description	6
1. Product Perspective	6
2. Problem Statement	6
3. Problem Solution	6
4. Proposed Methodology	6
5. Further Improvements	6
6. Data Required	6
7. Tools Used	7
8. Constraints	7
9. Assumptions	7
Design Details	8
1. Process Workflow	8
2. Error Handling	9
Performance	10
1. Reusability	10
2. Application compatibility	10
3. Resources Utilization	10
4. Deployment	10
Conclusion	11
Reference	12

Abstract

Google Analytics Customer Revenue Prediction is used to predict how much G-Store customers will spend. Customer spending on their project is clearly shown in the reports. Google Analytics automatically enriches your data by bringing Google machine-learning expertise to bear on your dataset to predict the future behavior of your users. With predictive metrics, you learn more about your customers just by collecting structured event data. So our goal is to predict the revenue that is going to be generated by those potential customers in the near future. In simple words, we are given with the users' past data and transactions (when they logged into G-store)., so by using this data we need to predict the future revenue will be created by those customers.

Introduction

1. Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project

2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms which should be understandable to the administrators of the system.

3. Definition

We are given with the users past data and transactions (when they logged into G-store)., so by using this data we need to predict the future revenue will be created by those customers.

The terms used in the projects are:

- CSV - comma-separated values

It is a text file that has a specific format which allows data to be saved in a table structured format.

General Description

1. Product Perspective

Google Analytics automatically enriches your data by bringing Google machine-learning expertise to bear on your dataset to predict the future behavior of your users. With predictive metrics, you learn more about your customers just by collecting structured event data.

2. Problem Statement

Predict the total revenue per customer to establish a better marketing decision,

3. Problem Solution

Develop a predictive model using G-Store dataset to predict the total revenue for the customer that helps in better user of marketing budget and also interpret the most impacting element on the total revenue prediction using different model.

4. Further Improvement

The project can be extended by using AI / ML to make it more predictable for customer revenue to provide accurate Results or Reports.

5. Data Required

- User's device related features
- Geographic features
- Web traffic features
- Date of visit

6. Tools Used

- Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib are used to build the whole model. Visual Studio Code is used as IDE.
- For visualization of the plots, Matplotlib and Seaborn are used.
- GitHub is used as version control system.

7. Constraints

Google Analytics Customer Revenue Prediction should be appropriate and easy to predict.

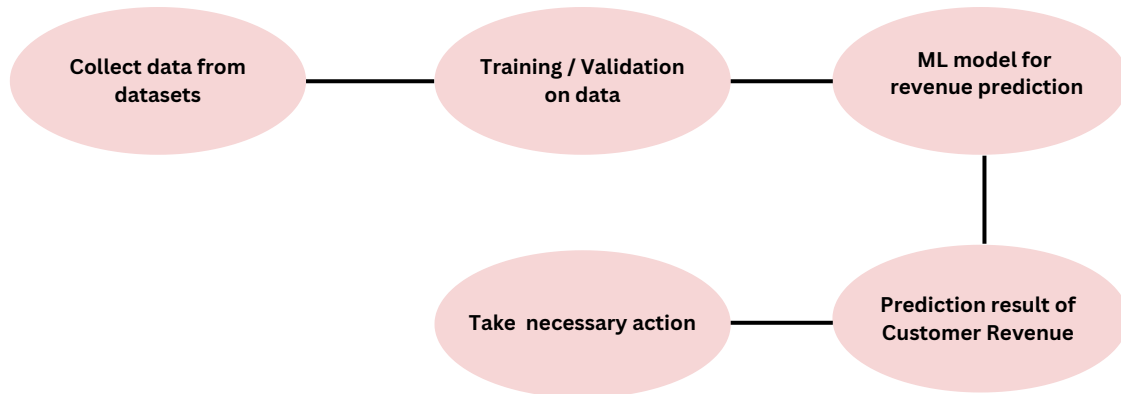
8. Assumptions

The main objective of the project is to collect and manage the datasets from various business resources to predict the revenue of the organization based on the training data by using Machine Learning. It is also assumed that all aspects of this project have the ability to make predictions as expected.

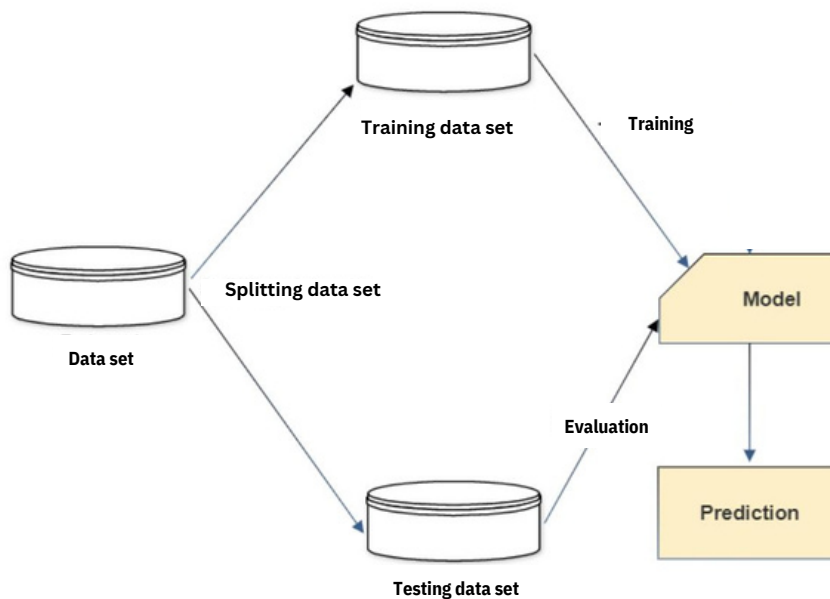
Design Details

1. Process Workflow

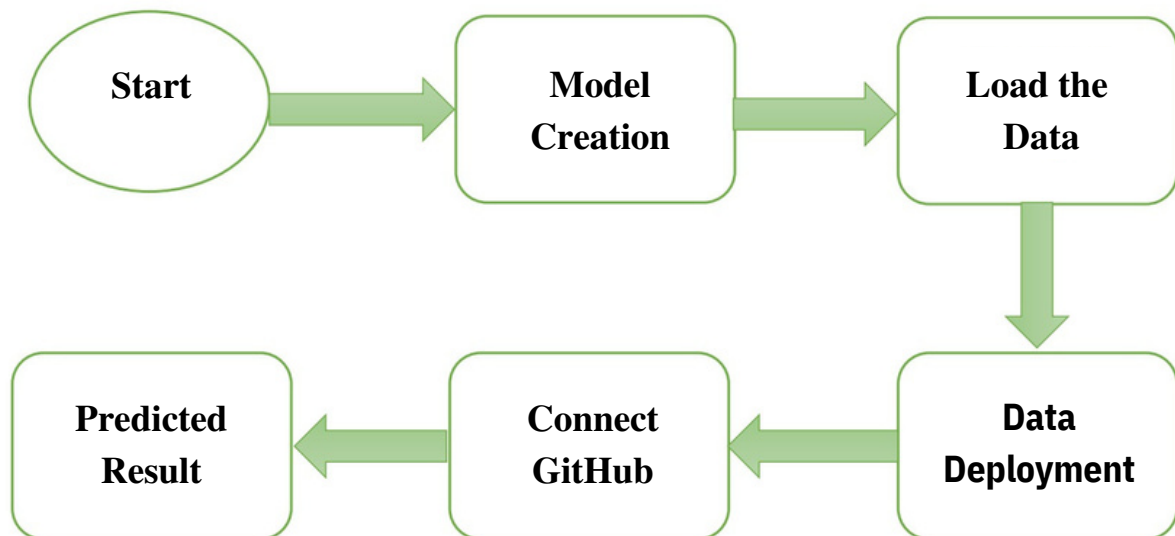
To predict customer revenue analytics we will use a machine learning model. Below is the process flow diagram.



Model Training and Evaluation



Deployment Process



2. Error Handling

The first problem we had was uploading our csv files with source code, but we were able to fix it. We then got error displaying the graph chart. All of the above errors have been resolved.

Performance

1. Reusability

Google Analytics Customer Revenue Prediction should be as accurate as possible, so that it will not mislead the user. Best possible model will be used to predict Customer revenue. Since we have used csv files and documentation is available, our project follows reusability.

2. Application compatibility

Since we are using python and it is compatible with any platform, we follow Application compatibility.

3. Resource utilization

At the initial stage, we were using high space to create the model. Once the model is created, our system only needs at least of 2GB RAM and 1 GB of storage to run the application smoothly. Whenever user tries to predict the Customer Revenue, system uses less than 10% of the processing power.

4. Deployment

The code is deployed in GitHub.

Conclusion

This project proposes the machine learning model for Google Analytics prediction. This model can be used for predicting revenue of customers. It will be helpful for us to manage and predict the datasets of the customers.

References

1. <https://www.kaggle.com/competitions/ga-customer-revenue-prediction/overview>.
2. https://scikit-learn.org/stable/user_guide.html
3. <https://numpy.org/doc/>
4. https://seaborn.pydata.org/examples/regression_marginals.html
5. https://seaborn.pydata.org/examples/scatterplot_matrix.html
6. <https://matplotlib.org/>
7. <https://pandas.pydata.org/docs/>