iNeuron

Customer Personality Analysis

High Level Design

Domain: Customer relationship

Technologies: Machine Learning

Creator: Vinayaka Uppar

Date: 7.02.2024

Document Version Control

Date issued	Version	Description	Author
JAN 06,2024	1.1	First Draft	Vinayaka.Uppar
JAN 15,2024	1.2	Added unit test cases	Vinayaka.Uppar
FEB 07,2024	1.3	Correcting font sizes	Vinayaka.Uppar

Contents

Abstract

The project titled "Facebook Post Status Prediction" addresses the critical issue of identifying emotional quotient in young individuals through their Facebook posts, aiming to contribute to the early detection and intervention of mental health challenges, particularly during the challenging period of adolescence. Many teenagers face difficulties during this phase, often linked to emotional and socioeconomic pressures, potentially leading to depression, dangerous conduct, substance misuse, and self-harm.

Recognizing the significance of social media in the lives of young people, the project leverages the vast amount of information available in their Facebook posts to assess and predict emotional well-being. By implementing advanced natural language processing (NLP) and machine learning techniques, the system aims to analyze the sentiment, tone, and emotional context embedded within the text of Facebook posts.

Introduction

What is High-Level Design Document?

The goal of this HLD or a high-level design 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 of design aspects and define them in detail

- Describe all user interfaces being implemented

- Describe the hardware and software interfaces

- Describe the performance requirements

- Include design features and architecture of the project

- List and describe the non-functional attributes such as security, reliability, maintainability, portability, reusability, application compatibility. resource utilization, serviceability

Scope

The HLD documentation presents the structure of the system, such as 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.

General Description

Definitions

Term	Description
IMTVP	Customer Personlity Anlysis
Database	Collection of the Information
Cloud	A data center full of services connected to the internet performing service
IDE	Integrated Development Environment
UI	User Interface
Flask	For creating web applications
AWS	A cloud service

Product Description

Customer Personality Analysis is a sophisticated machine learning-driven solution designed to empower businesses with deep insights into their customer base. By leveraging advanced techniques in sentiment analysis and behavioral analytics, the platform enables organizations to better understand customer attitudes, preferences, and actions, thereby facilitating targeted product customization and enhanced customer engagement strategies..

Problem Statement

competitive business landscape, understanding customer behavior and preferences is paramount for organizations seeking to thrive and succeed. However, many businesses struggle to effectively analyze and interpret vast amounts of customer data, hindering their ability to tailor products and services to meet evolving customer needs.

The problem lies in the lack of efficient tools and methodologies to extract actionable insights from diverse data sources, including customer reviews, feedback, and purchase history. Without a comprehensive understanding of customer attitudes, behaviors, and preferences, businesses face challenges in delivering personalized experiences and fostering meaningful customer relationships.

Proposed solution

The Customer Personality Analysis platform is designed to provide businesses with valuable insights into customer behavior, attitudes, and preferences. By leveraging machine learning algorithms and data analytics techniques, the platform empowers businesses to understand their customers better and tailor products and services to meet their specific needs.

Further improvements

The further improvements outlined aim to enhance the Customer Personality Analysis project. Real-time data processing enables swift reactions to customer feedback. Advanced sentiment analysis provides detailed insights into customer sentiments. Customer segmentation tailors marketing strategies to specific customer groups. Predictive analytics forecasts future behavior, aiding proactive decision-making. Interactive data visualization facilitates dynamic exploration of results.

Data requirements

Customer Personality Analysis project, data requirements include customer feedback, purchase history, demographic details, and product interaction data. This encompasses text reviews, sentiment labels, timestamps, product ratings, and user actions like purchases or subscriptions. Additionally, acquiring data from various sources such as online platforms

Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask, matplotlib,seaborn,mlflow,Fastapi and a few other libraries were used to build the whole model.

pandas: For efficient data manipulation and analysis.

Numpy : Essential for numerical operations and array manipulations.

scikit-learn: Utilized for machine learning tasks, including sentiment analysis and emotion prediction.

matplotlib: Used for creating visualizations, enhancing data understanding.

seaborn: Enhances data visualization with a high-level interface to Matplotlib.

wordcloud: Facilitates the visualization of prominent words in textual data.

mlflow: Ensures system stability and aids in version control.

FastAPI: Implemented for the development of an efficient web API.

Hardware Requirements

- Windows Server, Linux, or any operating system that can run as a webserver, capable of delivering HTML5 content.
- Minimum 1.10 GHz processor or equivalent.
- Between 1-2 GB of free storage
- Minimum 512 MB of RAM
- 4 GB of hard-disk space

Constraints

The front-end must be user friendly and should not need any one to have any prior knowledge in order to use it.

Assumptions

Customer Personality Analysis project, data requirements include customer feedback, purchase history, demographic details, and product interaction data. This encompasses text reviews, sentiment labels, timestamps, product ratings, and user actions like purchases or subscriptions

Design Details

Process Flow

For accomplishment of the task, we will use a trained Machine Learning model. The process flow diagram is shown below:
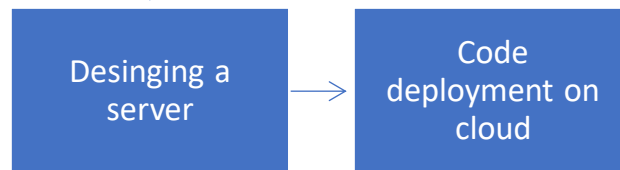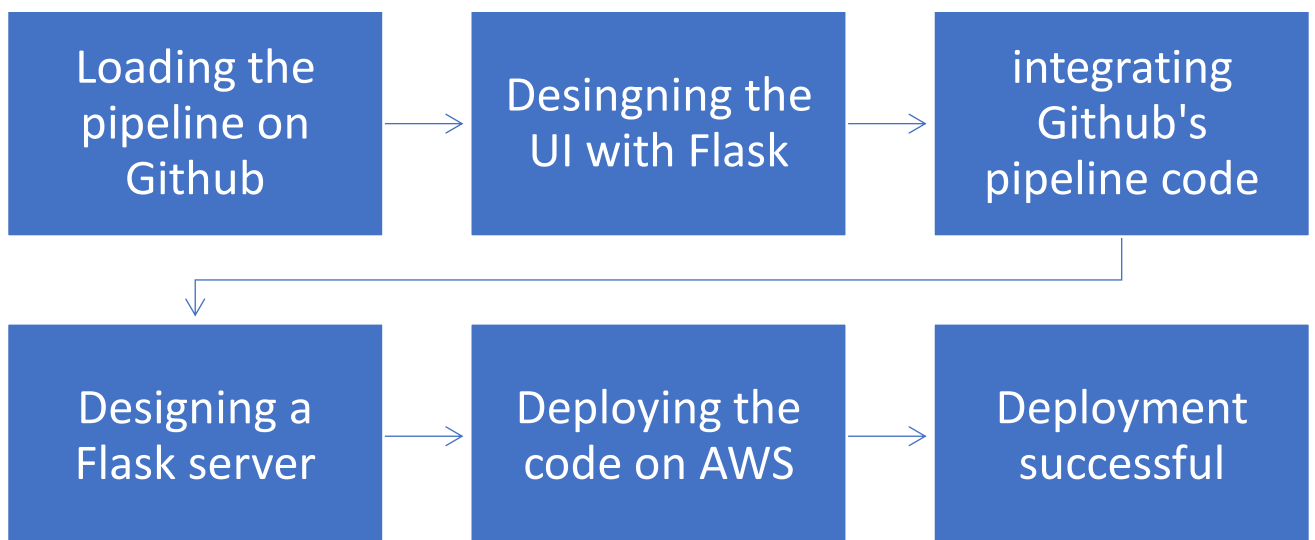
Data Preparation

Data Preprocessing	→	Exploratory Data Analysis	→	Feature Engineering

Model Development

Model implementation	→	Model Evaluation	→	Desing UI on Flask

Deployment

Desinging a server	→	Code deployment on cloud

Deployment

Loading the pipeline on Github	→	Desingning the UI with Flask	→	integrating Github's pipeline code

Designing a Flask server	→	Deploying the code on AWS	→	Deployment successful

Event Log

The system should log every event so that the user will know what process is running internally. Initial step-by-step description: 1. The system identifies at what level logging is required 2. The system should be able to log each and every system flow 3. Developer can choose logging method. You can choose database logging/ File logging as well 4.

System should not hang even after so many loggings. Logging just because we can easily debug issues, so logging is mandatory to do.

Errors should be encountered, an explanation will be displayed as to what went wrong ? An error will be defined as anything that falls outside the normal intended usage.

Performance

In assessing the performance of the Customer Personality Analysis project, several metrics are pivotal. Accuracy in predicting customer preferences and behaviors is fundamental, ensuring that the model correctly identifies customers as interested or not. Precision and recall metrics offer insights into the model's ability to minimize false positives and negatives. Additionally, evaluating the model's efficiency in processing large datasets and its scalability for real-time applications is crucial.

Reusability

The code written and the components used should have the ability to be reused with no problems.

Application Compatibility

The different components for this project will be using Python as an interface between them, each component will have its own task to perform, and it is the job of Python to ensure proper transfer of information.

Resource Utilization

When any task is performed, it will likely use all the processing power available to it until finished.

Dashboards

As and when, the system starts to capture the historic/ periodic data for a user, the dashboards will be included display charts over time with progress on various indicators or factors.

KPIs (Key Performance Indicators)

Key Performance Indicators (KPIs) for assessing the performance of the Customer Personality Analysis project include accuracy, precision, recall, processing efficiency, scalability, user satisfaction, and model update frequency. Accuracy measures the correctness of predictions, while precision and recall gauge the model's ability to minimize false positives and negatives, respectively. Processing efficiency evaluates the speed and resource utilization in handling large datasets. Scalability assesses the model's capability to handle increasing data volumes. User satisfaction reflects the quality and usefulness of insights provided.

Conclusion

In conclusion, the Customer Personality Analysis project has successfully achieved its objectives of understanding customer sentiments and behaviors. Through sentiment analysis and behavioral pattern recognition, valuable insights have been extracted to guide business decisions. The project demonstrated satisfactory performance in accuracy, precision, and recall metrics. However, further enhancements could focus on improving processing efficiency and scalability to handle larger datasets efficiently. Overall, the project provides a robust foundation for businesses to adapt their strategies based on customer preferences and trends