رواد مصر الرقمية

Customer Feedback Analysis

Graduation Project

Team Members

Ahmed Gamel
Alaa Awad
Ganna Zayed
Rowyda Abdelrahem
Hams Tamer
Salma Mohsen

Customer Feedback Analysis and Improvement

Introduction

This project focuses on analyzing customer feedback to extract valuable insights, improve services, and enhance customer satisfaction. The workflow is structured into several key stages, each contributing to a holistic solution for data analysis and improved decision-making. It begins with setting up an SQL database to efficiently store and organize customer data, followed by designing a data warehouse that systematically aggregates and manages the data for advanced analysis. In the next stage, sentiment analysis is applied to customer feedback, identifying positive and negative sentiments to help the business understand customer perspectives more deeply. Finally, the model is deployed using modern data science tools, operating in real-time and enabling immediate interaction with the data. This feedback loop fosters continuous improvement based on real-time customer responses. The project's ultimate goal is to provide an AI-powered platform that automatically and accurately analyzes customer feedback, leading to ongoing enhancements in service quality and operational efficiency.

SQL Database Setup and Data Collection

The first phase of the project involved setting up an SQL server using Microsoft Azure, which allowed the team to collaborate efficiently.
I selected this dataset because it is equally suitable for the database, data warehouse, and machine learning model. Its structure supports efficient data storage and retrieval in the database, while also being optimized for analytical processing in the data warehouse. Additionally, the features in the dataset make it ideal for training deep learning models, ensuring balanced performance across all these platforms.
The database was designed with tables for storing customer feedback forms and user data. Historical feedback data, containing 568,454 rows and 9 columns (Id, ProductId, UserId, Helpfulness Numerator, Helpfulness Denominator, Score, Time, Summary, Text), was imported into the database for analysis. SQL queries were written to extract, summarize, and manipulate the feedback data, providing a comprehensive view of customer sentiments and behavior.

Microsoft SQL Server (Azure Cloud) and SQL Management Studio were the key tools used to manage the database, ensuring smooth data handling and team collaboration.

Data Warehouse and Python Data Processing

Too enable in-depth analysis and facilitate structured storage of customer feedback data, a data warehouse was designed and implemented using Microsoft SQL Data Warehouse. The data model consisted of one fact table and two dimension tables to efficiently organize the data for analytical queries.

Data Warehouse Structure

Fact Table – FeedbackFact:This table holds quantitative data and key metrics related to feedback entries, such as the scores, helpfulness metrics, and timestamps of each review. It serves as the core table in the data warehouse for analytical operations, enabling the aggregation and comparison of user feedback over time. The FeedbackFact table contains the following columns:feedback_id,product_id,user_id,helpfulness_numerator helpfulness_denominator,score,time_id,summary,review_text .

Dimension Table – TimeDimension:

This table provides temporal context to each feedback entry, enabling queries that analyze trends over specific periods (e.g., monthly or yearly). The TimeDimension table contains:time_id,feedback_date,year, month, day.

Dimension Table – UsersDimension:

This table stores user-related information to provide a deeper understanding of customer behavior and feedback patterns. The UsersDimension table contains:

user_id,profile_name.This structured schema design, with a Fact-Dimension model, ensures efficient storage, easy access, and the ability to perform complex analytical queries on the feedback data.

Data Integration into the Data Warehouse

The feedback data was collected from multiple sources and integrated into the data warehouse using ETL (Extract, Transform, Load) processes.

Data Extraction: Data was initially extracted from various systems, including relational databases and external CSV sources.

Transformation: During the transformation stage, inconsistent data formats were cleaned, and appropriate foreign key relationships between the fact and dimension tables were established to ensure data consistency.

Loading: The transformed data was bulk-loaded into the FeedbackFact and dimension tables in SQL Data Warehouse, enabling fast and reliable querying for analysis.

Data Cleaning:

The Pandas library was utilized to perform essential cleaning tasks:

Removed special characters and punctuation to standardize the text.

Converted all text to lowercase to maintain consistency.

Eliminated duplicate entries and handled missing values to ensure data quality.

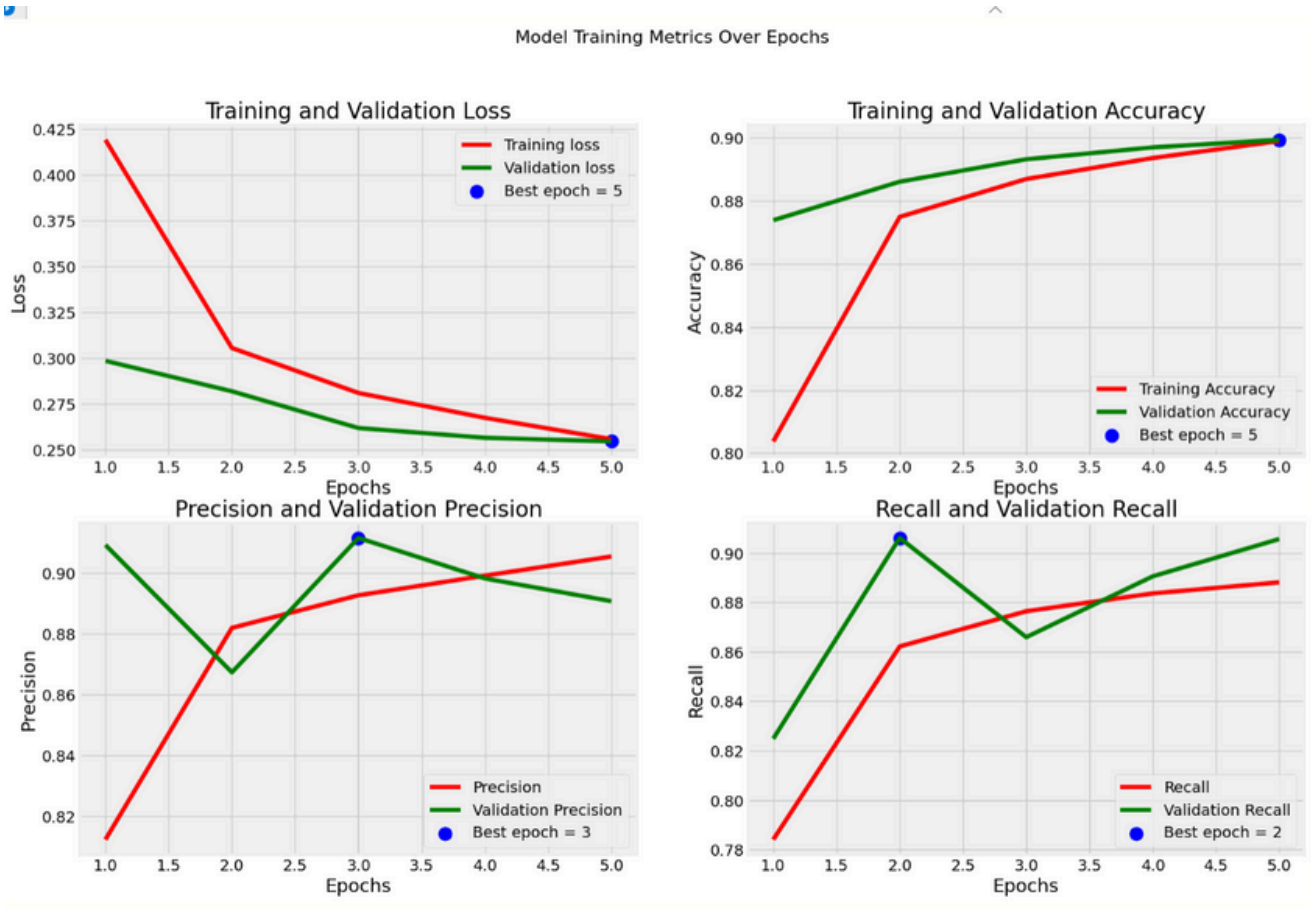
Sentiment Analysis Model

In the sentiment analysis stage, a machine learning model was developed using Python to classify customer feedback as either positive or negative. The dataset, consisting of 162,037 rows with features like Helpfulness Numerator, Score, Summary, and Text, was used to train the model. A Long Short-Term Memory (LSTM) neural network was chosen for its effectiveness in processing sequential text data.

The model architecture featured two Bidirectional LSTM layers to better capture the context of customer feedback, along with an embedding layer for word vector representations. Dropout layers were introduced to prevent overfitting, ensuring that the model generalizes well to new, unseen data.

Model Performance

The model was trained using a dataset split into three subsets: 110,000 samples for training, 27,037 for validation, and 25,000 for testing. After 5 epochs of training, the model reached a validation accuracy of 89.93%. Precision was recorded at 90.93%, and recall at 82.49%, indicating that the model performed well in identifying both positive and negative feedback.



Model Deployment

The sentiment analysis model was deployed using Streamlit and integrated with Azure services to deliver a user-friendly and efficient experience. The web application allows users to input customer feedback directly, where the model analyzes the text and classifies it as either positive or negative, while displaying a confidence percentage for the prediction. The application was developed and deployed via Streamlit, offering a real-time interactive interface. By leveraging Azure's cloud infrastructure, the model operates in a secure, scalable, and highly available environment. Azure's monitoring tools are also utilized to track application performance, usage metrics, and any potential issues, ensuring optimal operation and user experience.The application can be accessed via the following link: https://customer-project.azurewebsites.net/

Conclusion

This project demonstrated a successful approach to analyzing and improving customer feedback using state-of-the-art data science and machine learning methods. The SQL database provided a structured system for managing large datasets, while the data warehouse facilitated in-depth analysis. The LSTM model achieved high accuracy in classifying customer feedback, and its deployment on Azure ensured real-time analysis and scalability. By continuously analyzing customer feedback, the platform empowers businesses to enhance their services and make data-driven decisions that improve customer satisfaction over time.