

***Graduation Project Report***

***AI & Data Science Track***

**Tourism Sentiment Analysis**

**and**

**Recommendation System**

| **Team member** | **Contact** |
| --- | --- |
| **Aya Ehab** | [**aya\_eh\_ramadan@students.kasralainy.edu.eg**](mailto:aya_eh_ramadan@students.kasralainy.edu.eg) |
| **Mohamed Mahmoud** | [**mlpfa10a@gmail.com**](mailto:mlpfa10a@gmail.com) |
| **Maryam Sayed** | [**2021170506@cis.asu.edu.eg**](http://2021170506@cis.asu.edu.eg) |
| **Naden Mohamed** | [**nm1599481@gmail.com**](mailto:nm1599481@gmail.com) |
| **Ahmed Walied** | [**awalied2003239@gmail.com**](mailto:awalied2003239@gmail.com) |
| **Mosap Mohamed** | [**abdelghanymosap@gmail.com**](mailto:abdelghanymosap@gmail.com) |

***Table of Contents***

[1. Project Overview 3](#_fww8bdxu9uuv)

[2. Scope and Objectives](#_j4jd0a8l0tif) 4

[3. Methodology](#_1bdlj58drqck) 5

[a)](#_64n0dybtim1p) Data Collection5

[b) Data Preprocessing](#_hbbctmmtxf42) 5

[c)](#_lvd5wg7lvg2c) Sentiment Annotation6

[d)](#_pqdgtuvwk3ry) Handling Class Imbalance [6](#_pqdgtuvwk3ry)

[e)](#_qf91iy30i4qb) Text Vectorization with TF-IDF and Label Encoding [6](#_qf91iy30i4qb)

f ) Model Training & Evaluation……………………………7

[4. System Architecture & Deployment](#_n7z5h1771epn) 8

[5. Results and Validation](#_64h366n0xfq1) 9

[6. Challenges and Solutions 9](#_aea8d72wom8p)

[7. Future Work 1](#_ksn9ffoj1xkv)1

[8. Conclusion 1](#_lozrpud7hqfo)2

9. Appendix ………………………………………………..….13

## 1. [Project Overview](#_fww8bdxu9uuv)

Tourism is a vital pillar of Egypt's economy, contributing significantly to GDP, employment, and international perception. With its ancient monuments, diverse landscapes, and vibrant culture, Egypt attracts millions of visitors annually. However, despite its global appeal, recurring challenges such as inconsistent service quality, overcrowding, and infrastructure limitations often result in negative visitor experiences that can harm the country’s reputation and tourism sustainability.

This project explores how Natural Language Processing (NLP) and Machine Learning (ML) can be used to extract meaningful insights from tourist-generated content such as online reviews, social media posts, and travel blogs. By analyzing this unstructured textual data, we aim to understand visitor sentiments, detect common pain points, and offer practical recommendations for service improvement.

We collected over 30,000 reviews from Tripadvisor and Quora, covering major attractions across Egypt. The reviews were cleaned, labeled, and analyzed using traditional machine learning algorithms (Logistic Regression, Naive Bayes) and deep learning models (LSTM, GRU). Additionally, a large language model (LLM) was integrated to generate recommendations from negative reviews, and a web application was built using Streamlit for real-time sentiment prediction.

This project provides a practical tool that tourism stakeholders—including business owners, tourism boards, and policymakers—can use to monitor feedback, identify systemic issues, and take data-driven actions to enhance the overall tourist experience in Egypt.

# 2. Scope and Objectives

# **Scope**

This project focuses on analyzing sentiment from English-language online reviews related to tourism in Egypt. The reviews were primarily collected from Tripadvisor and span various tourist attractions, including historical sites, museums, natural landmarks, and cultural venues. The scope of the project includes:

* Web scraping and data collection from Tripadvisor.
* Text preprocessing and manual annotation for sentiment classification.
* Training and evaluating multiple machine learning and deep learning models to classify reviews as positive or negative.
* Integration of a free, open-source large language model (LLM) to extract actionable insights from negative reviews.
* Development of a Streamlit-based web application that allows users to enter a review and receive both sentiment prediction and recommendations.
* Visualization dashboard for monitoring trends and keyword analysis using Dash**.**

### **Objectives**

* Sentiment Classification: Build an accurate model to classify tourist reviews into positive or negative sentiment using NLP techniques.
* Deep Learning Integration: Evaluate LSTM and GRU models for handling long-form textual data effectively.
* Insight Generation: Use an LLM to analyze negative feedback and generate potential improvement suggestions.
* Interactive Deployment: Create an accessible Streamlit web interface where users can input reviews and receive predictions and recommendations.
* Stakeholder Empowerment: Provide tools that enable tourism professionals to monitor tourist feedback, identify recurring issues.

# 3. Methodology

## Data Collection

Tourism reviews were collected from publicly available online sources like TripAdvisor and Quora using a web scraping pipeline.

## Data Preprocessing

The reviews underwent several preprocessing steps:

* Lowercasing, URL and punctuation removal
* Emoji conversion to text
* Slang replacement via a custom slang dictionary
* Contraction expansion
* Tokenization, stopword removal, and lemmatization using NLTK
* Texts were then tokenized using Tokenizer() and padded for model input.

## Sentiment Annotation

we used ensemble of pretrained sentiment analysis models was employed:

* Binary Transformer Model
* VADER Sentiment
* Three-Class Transformer Model
* Ensemble Labelingfor the final sentiment using majority voting across the three models.

## Handling Class Imbalance

The original dataset was heavily imbalanced, with over 34,000 positive reviews and only ~4,000 negative ones. To avoid training bias:

* The dataset was split first into training and test sets.
* Upsampling was applied only to the training set duplicating minority (negative) class samples.
* The test set was left imbalanced for realistic performance evaluation.

## Text Vectorization with TF-IDF and Label Encoding

To convert text into numerical features for classical models:

* Used TfidfVectorizer to compute **TF-IDF scores**.
* Limited vocabulary to the **top 5,000 most important words**

To prepare the target variable for training:

* Used LabelEncoder to convert string sentiments into integers:  
  'negative' → 0  
  'positive' → 1

**f ) Model Training & Evaluation**

| **Model** | **Type** | **Accuracy** |
| --- | --- | --- |
| **Logistic Regression** | **Classical ML** | **87.27%** |
| **LGBMClassifier** | **Classical ML** | **86.8%** |
| **XGBoost Classifier** | **Classical ML** | **87%** |
| **Support Vector Machine (SVM)** | **Classical ML** | **88.2%** |
| **Stochastic Gradient Descent (SGD)** | **Classical ML** | **87.26%** |
| **Random Forest** | **Classical ML** | **91.12%** |
| **LSTM (Long Short-Term Memory)** | **Deep Learning (RNN)** | **90.09%** |
| **GRU (Gated Recurrent Unit)** | **Deep Learning (RNN)** | **91%(Best)** |

# 4. System Architecture & Deployment:

To make the sentiment analysis system accessible and user-friendly, we developed a web application using **Streamlit**, a lightweight Python framework for creating interactive data apps. The deployed system consists of several integrated components designed for real-time review classification and actionable feedback generation.

#### Architecture Overview

The system has the following core components:

| **component** | **Description** |
| --- | --- |
| app.py | The main Streamlit script responsible for running the web app interface, capturing user input, performing preprocessing, and displaying outputs. |
| gru\_model.h5 | The trained deep learning model (GRU-based) used for sentiment prediction. |
| tokenizer.pkl & label\_encoder.pkl | Preprocessing tools used to convert raw input into the format required by the GRU model. |
| LLM Model | A module that interacts with a Large Language Model (LLM) to generate tourism improvement recommendations for negative reviews. |

#### LLM Integration for Recommendations

* Initially, we experimented with **Ollama** open-source LLMs, but due to memory overhead and performance limitations on local machines, it was not feasible for production.
* The solution was migrated to **Gemini Flash (via Langchain and Google Generative AI)** for generating concise, tailored suggestions based on negative reviews.
* The integration uses langchain\_google\_genai.ChatGoogleGenerativeAI, and prompt engineering was applied to generate realistic tourism management insights.

#### Deployment

we used Streamlit to deploy the app

#### Output Functionality

* The app predicts the sentiment of a user-input review.
* For **negative** reviews, the app automatically sends a prompt to Gemini Flash and displays **recommendations** for improving the tourist experience.
* For **positive** reviews, the app displays only the sentiment classification.

# 5.Results and Validation

### **Results and Discussion**

This section presents the performance outcomes of various machine learning and deep learning models trained on labeled tourism review data, followed by an interpretation of their implications for real-world deployment.

#### Model Choose

The GRU (Gated Recurrent Unit) model outperformed all other architectures, achieving the highest overall scores.

#### Sentiment Distribution and Predictions

Model predictions showed a fairly balanced classification of both positive and negative reviews when tested on the real-world TripAdvisor dataset. This indicates the success of earlier class-balancing and vectorization strategies.

#### LLM-Based Recommendation Integration

To enrich the insights, we integrated Google’s **Gemini 2.0 Flash** LLM. It was prompted only when the sentiment prediction was **negative**, using contextual prompt engineering to suggest actionable tourism improvements. This step provided interpretability and enhanced the system's practical utility.

**6.Challenges and Solutions**

During the development of this tourism sentiment analysis system, we encountered several key challenges, each of which required careful problem-solving:

#### 1. Inaccurate Annotation from Ratings

Initially, we relied on user ratings from TripAdvisor to label sentiments automatically. However, we discovered that many ratings did not align with the actual sentiment expressed in the review text, leading to noisy labels and reduced model accuracy.

**Solution:** We replaced this heuristic with automated labeling using multiple sentiment analysis models, including transformer-based binary classifiers (RoBERTa), VADER, and a three-class model. These predictions were combined using an ensemble voting approach to improve labeling consistency. This strategy eliminated dependency on subjective ratings and enabled the incorporation of high-quality external data sources, such as Quora tourism-related discussions.

#### 2. Difficulty Integrating Multi-Source Data

Integrating datasets from TripAdvisor and Quora initially introduced challenges due to format differences and inconsistent labeling. The earlier dependency on TripAdvisor’s rating system also limited the inclusion of diverse sources.

**Solution:** By automating the annotation process using model ensembles, we decoupled our dataset from rating-based labels. This allowed seamless integration of multiple data sources after unifying preprocessing steps, enhancing the dataset's richness and diversity.

#### 3. Class Imbalance

The original dataset was heavily skewed toward positive reviews (over 34,000), with significantly fewer negative samples (about 4,000). This imbalance threatened to bias the model toward the majority class.

**Solution:** We addressed this by splitting the dataset into training and test sets, then applying **upsampling** only to the training set using sklearn.utils.resample. This preserved the natural imbalance in the test set for realistic evaluation, while ensuring balanced training.

#### 4. Recommendation system using Aspect-Based Sentiment Analysis (ABSA) Limitations

We initially explored using ABSA techniques to extract specific topics and sentiments from negative reviews (e.g., "cleanliness," "service"). However, existing models failed to group semantically related issues accurately, often missing nuances in feedback or misclassifying aspects.

**Solution:** We transitioned to using a large language model (LLM)—specifically **Gemini 2.0 Flash**—which enabled contextual understanding of full review texts. By providing a carefully crafted prompt, the model could identify pain points and suggest actionable recommendations with significantly improved accuracy and relevance.

#### 5. Local LLM Deployment Limitations

We experimented with using **Mistral** via **Ollama** for running LLMs locally. However, the model was large and required GPU resources.

**Solution:** We switched to using Gemini via API for efficient cloud inference. This solution was lightweight, scalable, and well-suited for deployment in environments like Streamlit, providing quick and reliable recommendation generation.

# 7.Future Work

#### 1. Multi-Language Support

Currently, the system primarily handles English reviews. Extending support to other commonly spoken languages (e.g., Arabic, French, German) would make the system more inclusive and representative of global tourists.

#### 2. Improved Aspect-Based Sentiment Extraction

Although the current LLM-based recommendation system performs well, fine-tuned aspect extraction models could be incorporated to pinpoint sentiment tied to specific aspects (e.g., "cleanliness", "pricing", "staff behavior") for more targeted improvements.

#### 3. Real-Time Sentiment Monitoring

Integrating live review streams from platforms like TripAdvisor, Google Maps, or Twitter through APIs could enable **real-time dashboard monitoring** for tourism stakeholders, allowing faster responses to emerging issues.

#### 4. Geospatial Analysis

Incorporating geographic data could enable **sentiment heatmaps**, helping stakeholders visualize which areas or attractions receive the most positive or negative feedback geographically.

# 8.Conclusion

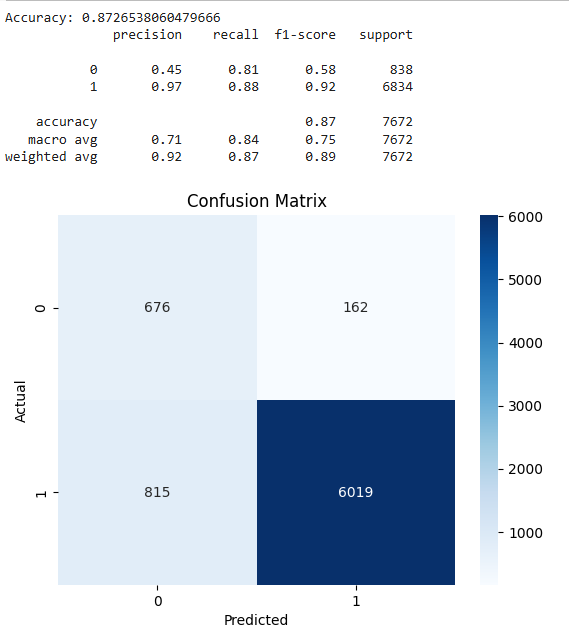
This project presents an end-to-end sentiment analysis system tailored for the tourism sector in Egypt, leveraging natural language processing (NLP), machine learning, and large language models (LLMs). By analyzing online reviews from platforms like TripAdvisor and Quora, we were able to classify sentiments, detect recurring issues in visitor experiences, and generate actionable recommendations.

To overcome noisy labels and rating inconsistencies, we automated the annotation process using multiple sentiment models and ensemble voting. We addressed class imbalance using strategic upsampling and trained several machine learning models, with deep learning models like GRU achieving the best performance. Additionally, a recommendation system powered by a generative LLM was integrated to suggest practical solutions based on negative reviews.

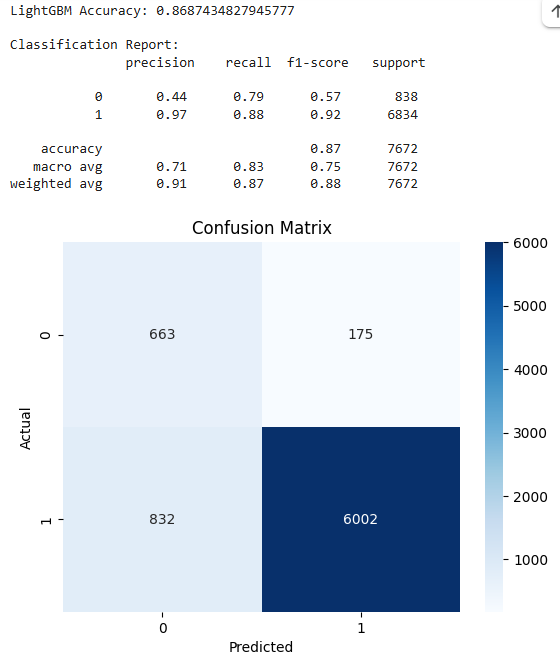
The system was deployed as an interactive Streamlit app, offering an accessible tool for tourism authorities and businesses to monitor feedback, understand visitor sentiment, and drive improvements in service quality. This solution lays the groundwork for a more responsive and data-driven tourism ecosystem in Egypt.

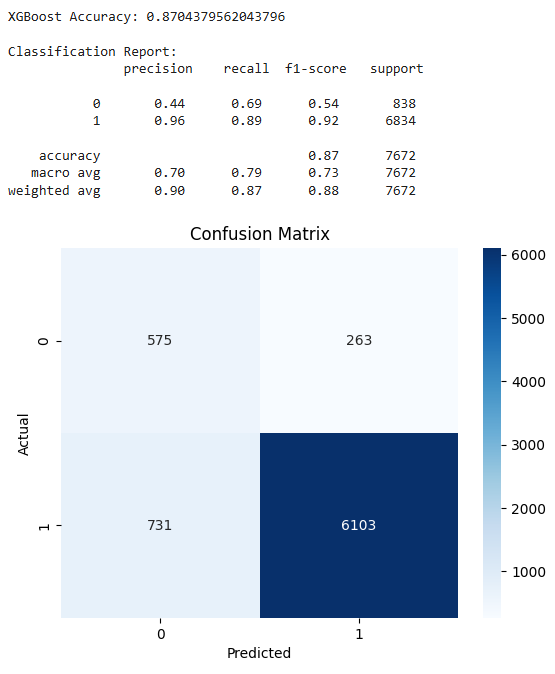
**9.Appendix**

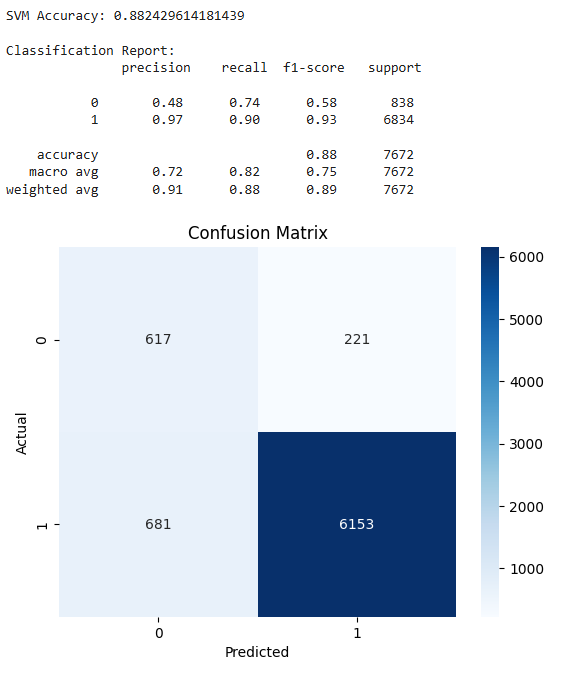
1. ***Model Evaluation***
2. **Logistic Regression**



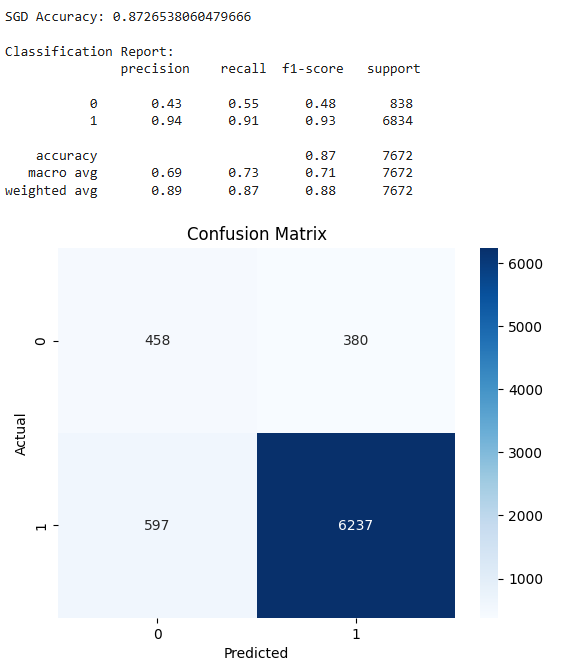
1. **LGBMClassifier**



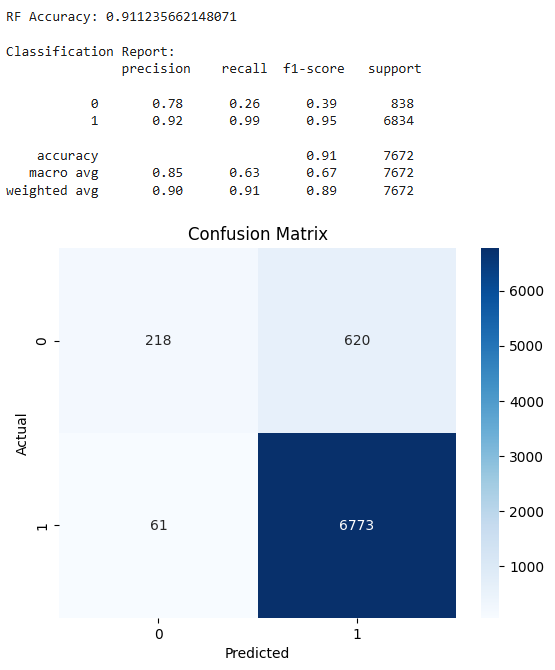
1. **XGBoost Classifier**
2. **SVM Classifier**

****

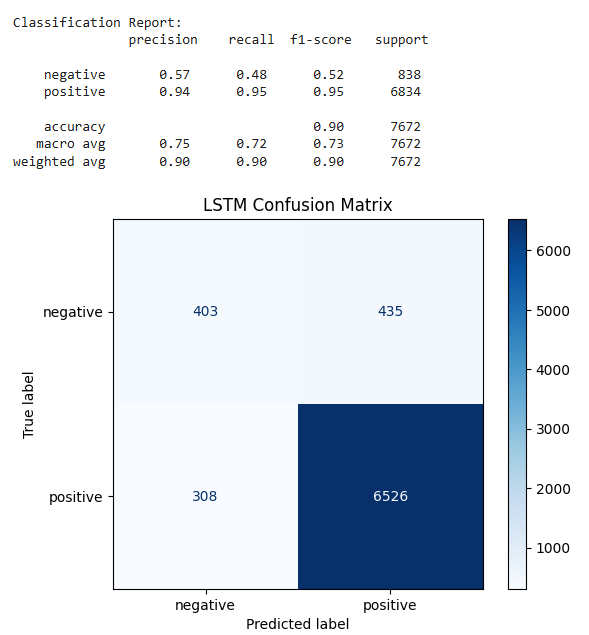
1. **SGD Classifier**

****

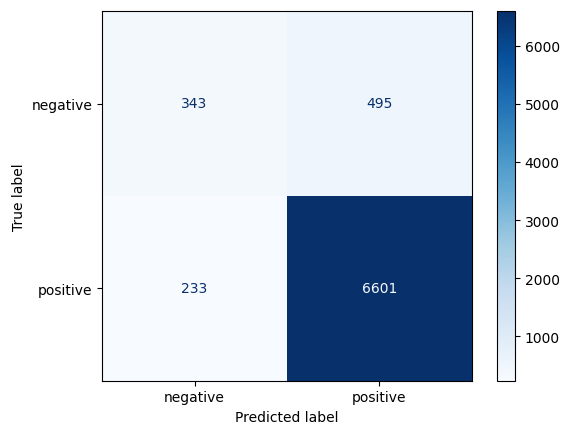
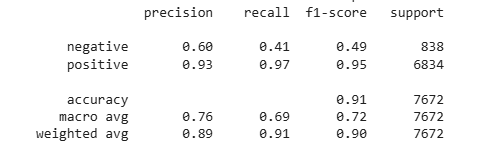
1. **Random Forest Classifier**

****

1. **LSTM**

****

1. **GRU**

****