TasteLens AI Team

Sentiment Analysis

Welcome to **TasteLens AI**'s Sentiment Analysis Project

What our prokect focuses on:
Our project focuses on developing a robust sentiment analysis system specifically for restaurant reviews.
By leveraging machine learning techniques, we aim to classify customer feedback into positive or negative sentiments accurately.
This analysis provides valuable insights for restaurants to improve their services, address customer concerns, and enhance overall satisfaction.

What We Aim to Achieve with This Presentation:

- Share the challenges of sentiment analysis in the restaurant industry.
- Highlight our methodology and innovative approach.
- Present the results of testing five machine learning models.
- Conclude with key takeaways and recommendations for future work.

Agenda

01

Problem Statement

02

Related Work

03

Proposed Methodology

Agenda

04

Results

05

Coclusion

1-Problem Statement.

The Challenge of Interpreting Restaurant Reviews

The Problem:
- Customer satisfaction is critical for restaurants, but understanding feedback from large volumes of unstructured data is challenging.
- Many existing tools are ineffective at interpreting nuanced language, such as sarcasm or mixed sentiments.

Key Issues:
- Inability to handle industry-specific jargon or restaurant-related terms.
- Misinterpretation of sarcastic or emotional reviews.
- Limited actionable insights for improving customer experience.

Our Objective:
- Build a robust sentiment analysis model that captures these nuances and classifies reviews accurately.

2-Related Work.

Learning from Existing Research

What We Studied:

1-Twitter sentiment analysis: Capturing sentiment from integrated resort tweets

International Journal of Hospitality Management 55 (2016) 16–24

Contents lists available at ScienceDirect

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhosman

Twitter sentiment analysis: Capturing sentiment from integrated resort tweets

Kahlil Philander[a, 1], YunYing Zhong[b,*]

[a] William F. Harrah College of Hotel Administration, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154, United States
[b] Hospitality Management, College of Health and Human Services, California State University, Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, United States

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 26 January 2016
Accepted 2 February 2016
Available online 11 March 2016

1. Introduction

With the rapid development of information and communication technologies, social networking sites and blogs have become an enormous repository for rich user generated content (UGC) (Thelwall et al., 2011). Information system scholars observed that business intelligence has evolved from analysis of traditionally structured business transaction data to studying unstructured and real-time data, driven by the development of online search engines, e-commerce, and social media (Chen et al., 2012). Among the existing collection of social media platforms, the Twitter microblog is one of the most popular mediums that have been adopted by both consumers and companies. Over half a billion Twitter messages (tweets) per day are being recorded (Kirilenko and Stepchenkova, 2014; Krikorian, 2013), and a recent customer engagement technology study reported that Twitter is among the top three most used social media platforms by US hotel and restaurant operators (Kim and Connolly, 2013).

The prevalence of Twitter and its prodigious amount of UGC creates important implications for the hospitality and tourism sector. Research has shown that online UGC has become an important information source that exerts critical influence on customers' brand perception, brand reputation, purchase decision making, and profitability (Browning et al., 2013; Leung et al., 2013; Vermeulen and Seegers, 2009; Ye et al., 2009, 2011; Zhang et al., 2010). Twitter may be particularly susceptible to the effects of electronic words of mouth (EWOM) due to its viral nature. According to Kwak et al. (2010) study, a tweet, if resent by a different user (retweeted), is expected to reach an average of 1000 registered users. In three years following Kwak et al.'s study, the total volume of daily tweets has grown over ten-fold (Krikorian, 2013; Weil, 2010).

While the viral influence of Twitter messages makes it valuable for firms to monitor and identify customer comments, the massive volume and high variance of real-time tweets has made that task extremely time consuming, costly, and often manually impossible (Chou et al., 2015; Claster et al., 2013). Even though firms have started to embrace Twitter as a marketing platform, because of these obstacles, they often fail to extract much objective quantitative feedback or establish competitive benchmarks (i.e. comparing firm performance in social media against competitor performance) from this available data.

Sentiment analysis, which is still a rather nascent arena, appears to be a promising tool in solving the above mentioned problem. This technique involves using algorithms to automatically extract and classify text data into sentiment categories: positive, negative and/or neutral (Chou et al., 2015; Pang and Lee, 2008). As compared to more traditional market research methods (e.g. surveys or opinion polls), sentiment analysis not only has the advantage of being more cost and time efficient in many cases, it is a non-intrusive method to extract consumers' opinions and sentiments in "real-time"—avoiding recall biases generally associated with post-consumption self-report measurements (Rylander et al., 1995). Furthermore, sentiment analysis can provide a temporal sentiment profile even on a second by second scale, which is not generally possible for survey based market analysis to achieve.

In this study, we demonstrate the application of sentiment analysis using Twitter data to build low-cost and real-time measures of hospitality customer attitudes/perceptions. Using a popular tourism destination (Las Vegas, NV) as a case study, we create a sentiment index for every Twitter account belonging to an

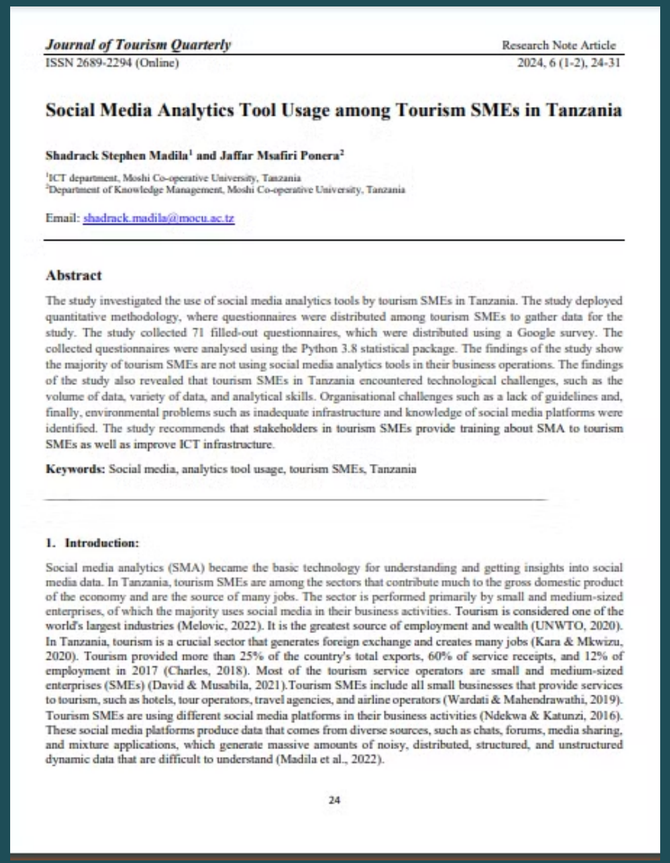
* Corresponding author. Tel.: +1562 985 4406.
E-mail addresses: kahlil.philander@unlv.edu (K. Philander), yunying.zhong@csulb.edu (Y. Zhong).
[1] Tel.: +1 702 895 1265.

http://dx.doi.org/10.1016/j.ijhm.2016.02.001
0278-4319/

Learning from Existing Research

What We Studied:

2- Analyzing Twitter to explore perceptions of Asian restaurants



Journal of Tourism Quarterly Research Note Article
ISSN 2689-2294 (Online) 2024, 6 (1-2), 24-31

Social Media Analytics Tool Usage among Tourism SMEs in Tanzania

Shadrack Stephen Madila[1] and Jaffar Msafiri Ponera[2]

[1]ICT department, Moshi Co-operative University, Tanzania
[2]Department of Knowledge Management, Moshi Co-operative University, Tanzania

Email: shadrack.madila@mocu.ac.tz

Abstract

The study investigated the use of social media analytics tools by tourism SMEs in Tanzania. The study deployed quantitative methodology, where questionnaires were distributed among tourism SMEs to gather data for the study. The study collected 71 filled-out questionnaires, which were distributed using a Google survey. The collected questionnaires were analysed using the Python 3.8 statistical package. The findings of the study show the majority of tourism SMEs are not using social media analytics tools in their business operations. The findings of the study also revealed that tourism SMEs in Tanzania encountered technological challenges, such as the volume of data, variety of data, and analytical skills. Organisational challenges such as a lack of guidelines and, finally, environmental problems such as inadequate infrastructure and knowledge of social media platforms were identified. The study recommends that stakeholders in tourism SMEs provide training about SMA to tourism SMEs as well as improve ICT infrastructure.

Keywords: Social media, analytics tool usage, tourism SMEs, Tanzania

1. Introduction:

Social media analytics (SMA) became the basic technology for understanding and getting insights into social media data. In Tanzania, tourism SMEs are among the sectors that contribute much to the gross domestic product of the economy and are the source of many jobs. The sector is performed primarily by small and medium-sized enterprises, of which the majority uses social media in their business activities. Tourism is considered one of the world's largest industries (Melovic, 2022). It is the greatest source of employment and wealth (UNWTO, 2020). In Tanzania, tourism is a crucial sector that generates foreign exchange and creates many jobs (Kara & Mkwizu, 2020). Tourism provided more than 25% of the country's total exports, 60% of service receipts, and 12% of employment in 2017 (Charles, 2018). Most of the tourism service operators are small and medium-sized enterprises (SMEs) (David & Musabila, 2021).Tourism SMEs include all small businesses that provide services to tourism, such as hotels, tour operators, travel agencies, and airline operators (Wardati & Mahendrawathi, 2019). Tourism SMEs are using different social media platforms in their business activities (Ndekwa & Katunzi, 2016). These social media platforms produce data that comes from diverse sources, such as chats, forums, media sharing, and mixture applications, which generate massive amounts of noisy, distributed, structured, and unstructured dynamic data that are difficult to understand (Madila et al., 2022).

Learning from Existing Research

What We Studied:

3- Sentiment Analysis of Restaurant Reviews Using Hybrid Classification Method

DUJE (Dicle University Journal of Engineering) 13:3 (2022) Sayfa 493-500

Dicle University
Journal of Engineering

Derleme Makalesi / Review Article

Farklı dillerdeki metinler üzerinde yapılan duygu analizi çalışmalarının incelenmesi

Analysis of sentiment analysis studies on texts in different languages

Seda Kılıçer[1,2*], Ruya Samli[2]

[1] Beykent Üniversitesi, Bilgisayar Mühendisliği Bölümü, sedakilicer@beykent.edu.tr
ORCID: 0000-0002-3464-6219
[2] İstanbul Üniversitesi-Cerrahpaşa, Bilgisayar Mühendisliği Bölümü, ruyasamli@istanbul.edu.tr
ORCID: 0000-0002-8723-1228

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 13 Temmuz 2022
Revizyon 24 Ağustos 2022
Kabul 15 Eylül 2022
Online 30 Eylül 2022

Anahtar Kelimeler:
Duygu analizi,
Sözlük tabanlı yaklaşım,
Makine öğrenmesi yaklaşımı,
Hibrit yaklaşım,
Sınıflandırma algoritmaları.

ÖZ

Teknolojinin gelişmesiyle birlikte veri miktarı hızlı bir şekilde artmaktadır, sürekli artan bu verilerden faydalı bilgilerin bakılmaksızın, ihtiyaçlarımız ile ilgili görüşlerin manuel bir şekilde verilere arasında uygun edilebilmesi çok mümkün olmayacaktır. Bunun için veriler işlenmediği sürece anlamlı analizlerin yapılarak faydalı bilgilerin elde edilmesi çok zor olacaktır. Bu sebeple bu büyük verilerin sistemler aracılığıyla hızlı bir şekilde bilgiye dönüştürülmesi ve insanlar tarafından kullanılabilir hale dönüştürülebilmesi hedeflenmektedir.
Duygu analizi, bir metnin olumlu, olumsuz veya nötr gibi duygulardan hangisine daha yakın olduğunun sonucunun farklı yöntemler kullanılarak elde edilmesidir. Bu çalışmada son on yılda (2013-2022) yapılan duygu analizi çalışmaları ele alınmış, değerlendirilmiş ve sınıflandırılmıştır. Bu amaçla sözlük tabanlı yaklaşım, makine öğrenmesi yaklaşımı ve hibrit yaklaşım yöntemleri kullanılarak duygu analizi yapılan 68 çalışma incelenmiştir. Çalışmalar ilk olarak dizeyi analizinde kullanılan yöntemlere göre, sonucunda verilerin elde edilme şekillerine ve incelenen metnin dilime göre sınıflandırılmıştır. Bunlar dışında duygu analizinde kullanılan yöntemlere göre uygulanan adımlar detaylı olarak incelenmiştir. Duygu analizi çalışmalarında elde edilen başarı oranlarının artırılabilmesi için incelenen çalışmalar tablolar aracılığıyla karşılaştırılmıştır. Her yöntem kendi içerisinde değerlendirilerek daha başarılı sonuçlar elde edilebilmesi için incelenen kriterler ve elde edilen sonuçlar arasındaki bağlantılar araştırılmıştır.

ARTICLE INFO

Article history:

Received 13 July 2022
Received in revised form 24 August 2022
Accepted 15 September 2022
Available online 30 September 2022

Keywords:
Sentiment analysis,
Dictionary-based approach,
Machine learning approach,
Hybrid approach,
Classification algorithms.

Doi: 10.24012/dumf.1143879

ABSTRACT

With the development of technology, the amount of data is increasing rapidly, it will not be possible to find useful information from this ever-increasing data and to determine the solutions related to our needs manually among the data. As long as the data is not processed for this purpose, it will be very difficult to obtain useful information by making meaningful analyzes. For this reason, it is aimed to transform this big data into information quickly through systems and to make it usable by people.
Sentiment analysis is the use of different methods to obtain the result of which of the emotions a text is closer to, such as positive, negative or neutral. In this study, sentiment analysis studies conducted in the last ten years (2013-2022) were discussed, evaluated and classified. For this purpose, 68 studies in which sentiment analysis was performed using dictionary-based approach, machine learning approach and hybrid approach methods were examined. Studies were first classified according to the methods used in sentiment analysis, and then according to the way the data were obtained and the language of the text examined. Apart from these, the steps applied according to the methods used in sentiment analysis were examined in detail. In order to increase the success rates obtained in sentiment analysis studies, the studies examined were compared through tables. Each method is evaluated within itself and the connections between the criteria examined and the results obtained are investigated in order to obtain more successful results.

* Sorumlu Yazar

Learning from Existing Research

Findings:

- Sentiment analysis in the restaurant industry has been explored but lacks precision in handling sarcasm and complex emotions.
- Most methods use basic models that don't fully utilize advanced machine learning techniques.

Gaps Identified:

- Absence of restaurant-specific preprocessing techniques.
- Over-reliance on general-purpose algorithms without optimization for sentiment nuances.

How We Address This:

- Introduced five advanced machine learning models and customized preprocessing for better results.

3-Proposed Methodology.

Flowchart

Data Collection
Searching for the best dataset to use

Data Preprocessing
Cleaning
Remove stop words
Feature Extraction
Make all Lowercase
Tokenization

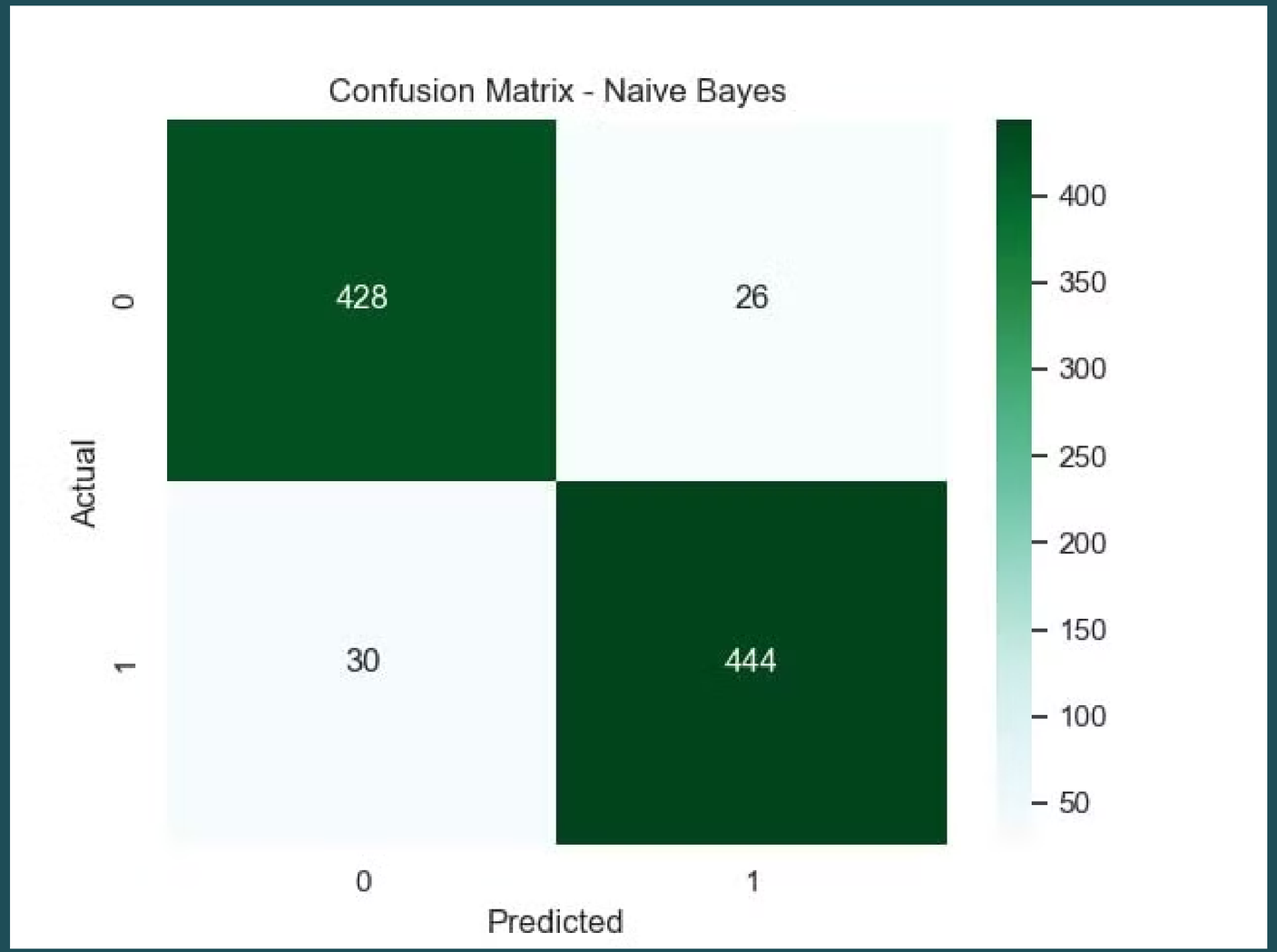
Model Training
Logistic Regression
Random Forest
SVM
Decision Tree
KNN
Naive Bayes

Model Evaluation
Split Data
Evaluation Metrics

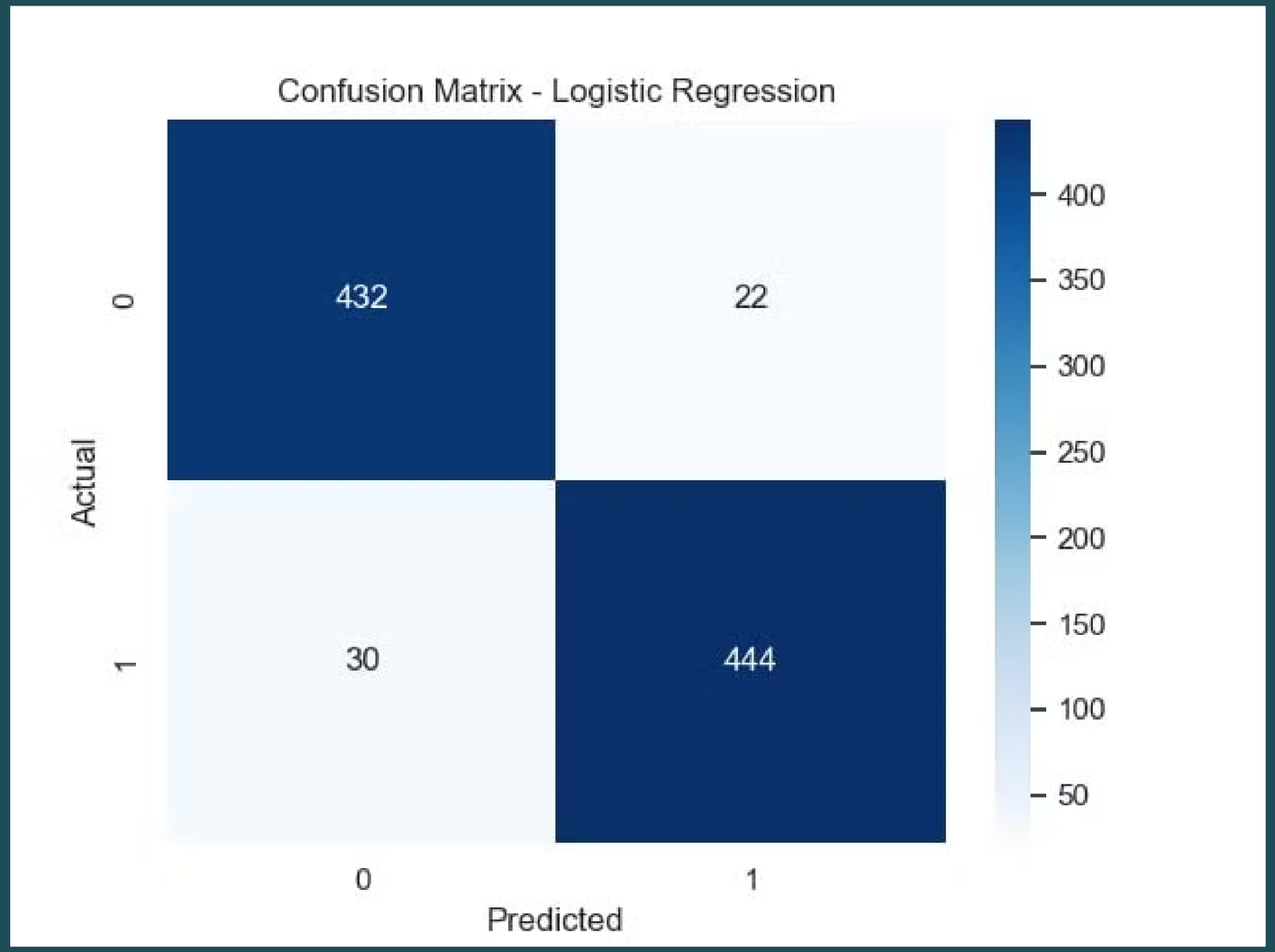
Results & Insights
Compare models and identify the best performer (Logistic).

Represent it in a simple GUI

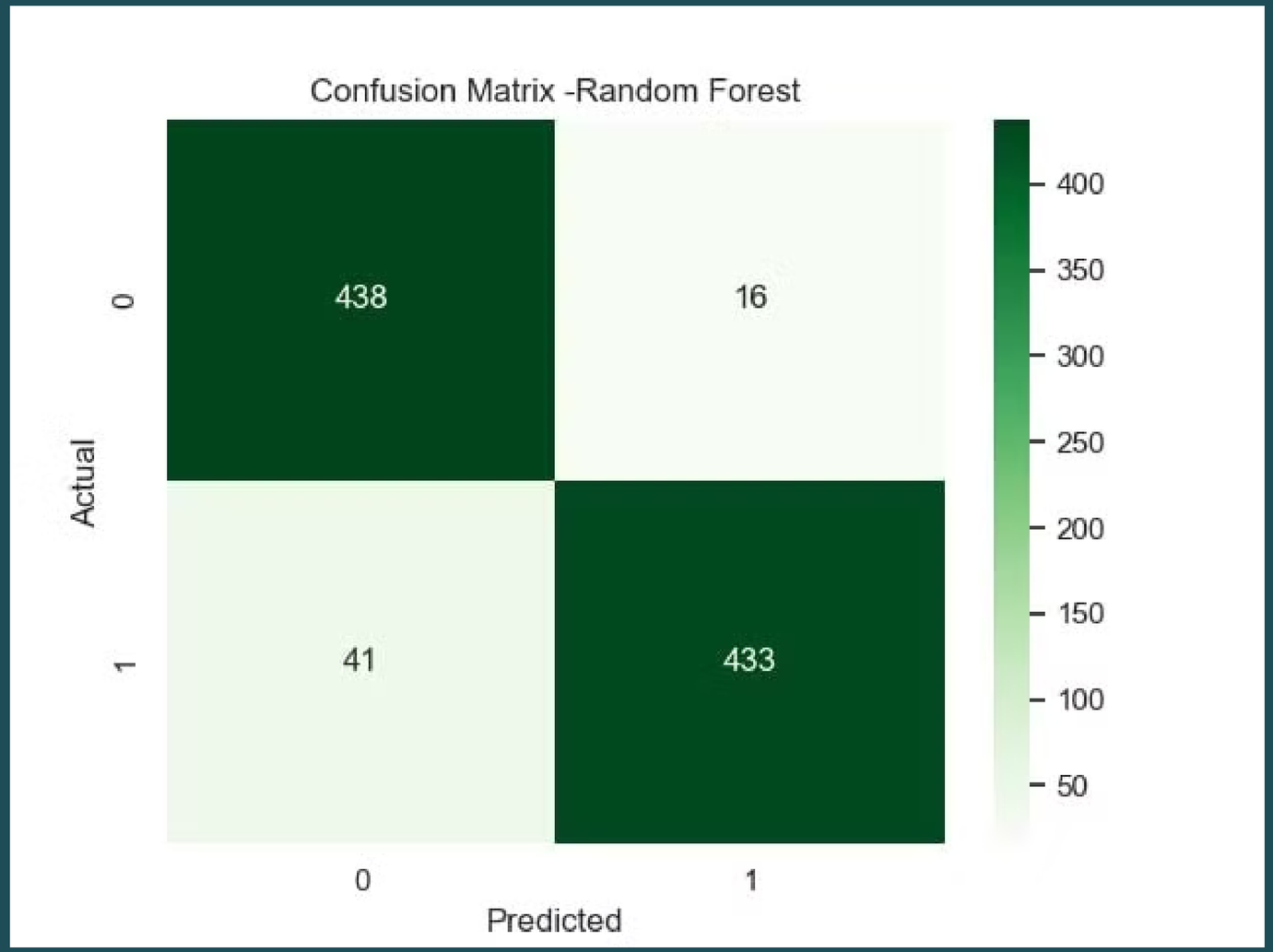
1- Confusion Matrix - Naive Bayes



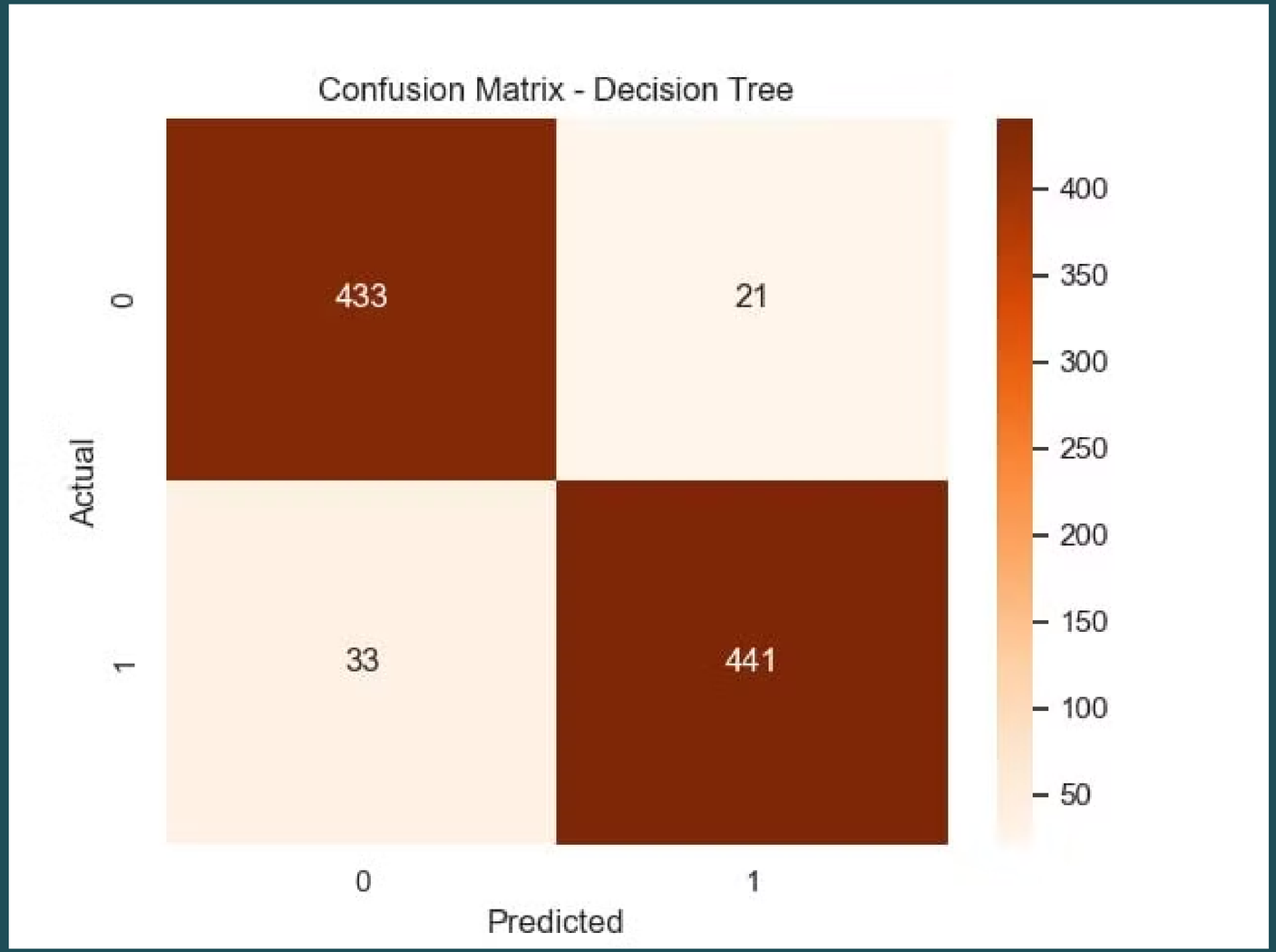
2- Confusion Matrix - Logistic Regression



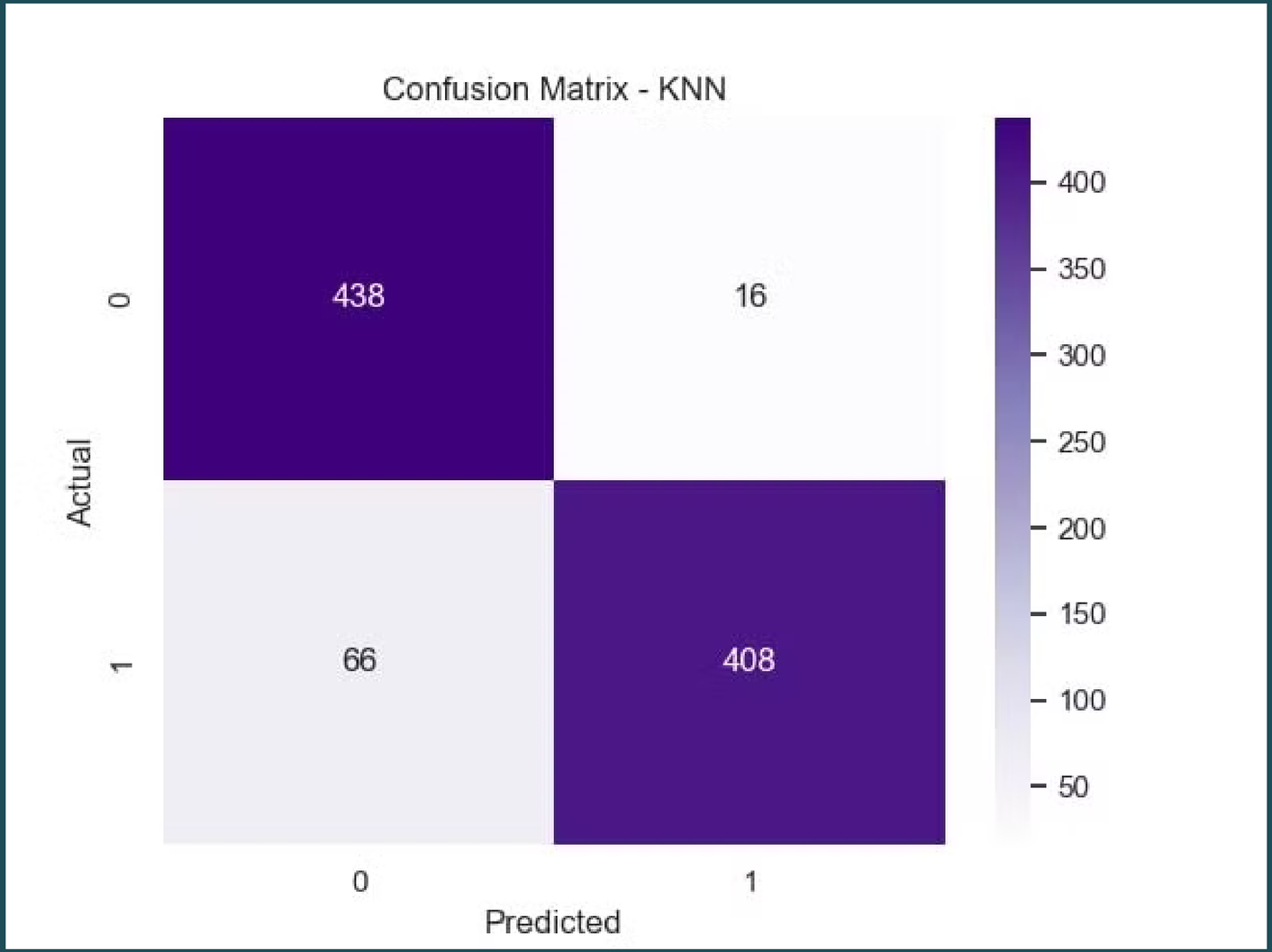
3- Confusion Matrix - Random Forest



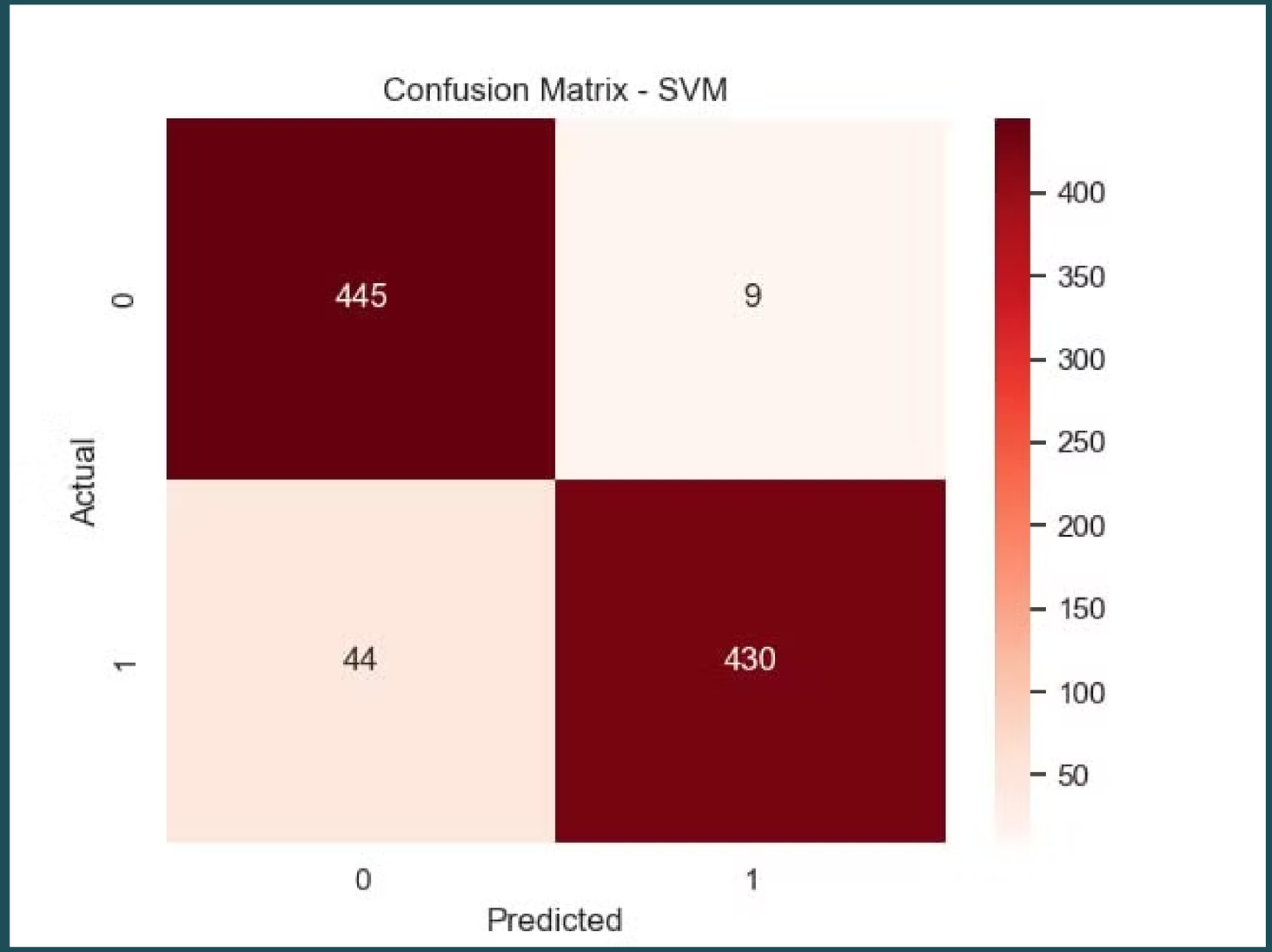
4- Confusion Matrix - Decision Tree



5- Confusion Matrix - KNN

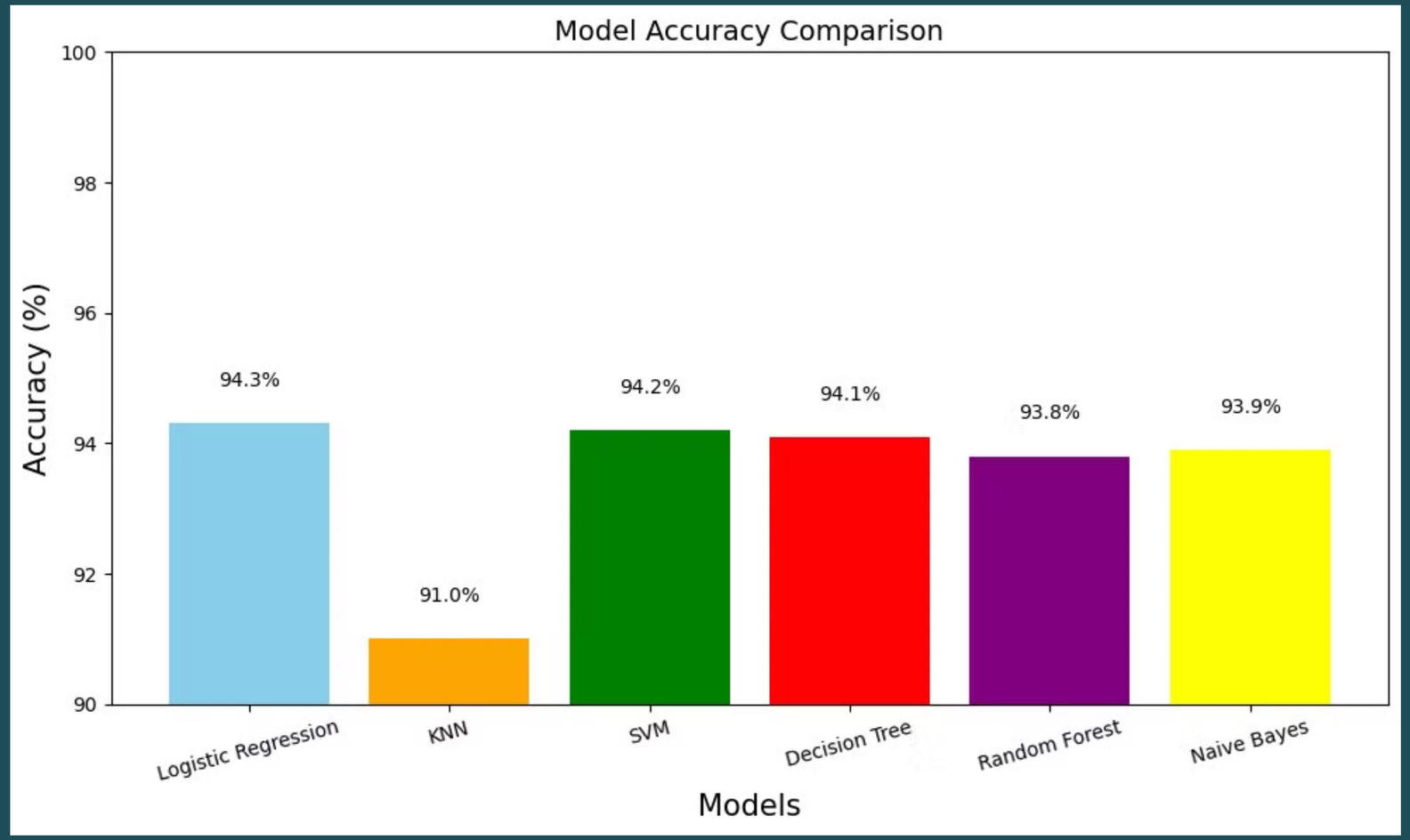


6- Confusion Matrix - SVM



4-Results.

Model Comparison Table



OUR GUI

So from the previous figure we found out the best model to use in pur project and on our data is the **Logistic Regression Model with a 94.3% Accuracy** , so we confirm it and made a simple GUI using streamlit library at python to make our project semi complete.

*it is a simple photos represent and clarify our gui

OUR GUI

It predicts the positive or negative review from the input

ex: Positive

OUR GUI

It predicts the positive or negative review from the input

ex: Negative

5-Conclusion.

Key Takeaways and Future Directions

- **Key Findings**:
  - Sentiment analysis provides actionable insights for improving customer experience.
  - Among the six models, **Logistic Regression** emerged as the most reliable and accurate for classifying sentiments.
  -
- **Contributions**:
  - Delivered a tailored machine learning solution for the restaurant industry.
  - Improved the classification accuracy by addressing domain-specific nuances.
  -
- **Future Work**:
  - Incorporate sarcasm detection to further enhance model accuracy.
  - Test the model on real-time restaurant tweets for live feedback analysis.
  - Expand the dataset to include more diverse restaurant types and customer demographics.

Thank You.

That's it, folks.

Try Pitch