FINAL PROJECT  
BUSINESS CASE

**Submitted by:**

Jijo John Joseph (200521777)

Table of Contents

1. Executive Summary 3

2. Introduction 3

3. Project Description 3

4. Technology Stack 4

5. Project Scope 4

6. Implementation Plan 5

7. Design Diagram 5

8. Risk Assessment 6

9. Cost Benefit Analysis 6

10. Deliverables 6

11. Conclusion 6

12. References and Citation 7

# 1. Executive Summary

This project aims to perform sentiment analysis on Yelp reviews using MongoDB, RapidMiner, and Flask. By analysing customer sentiment, we can gain insights into customer satisfaction and help make out the best ice cream and desert shops in Toronto.

# 2. Introduction

Yelp is a renowned platform where customers express their experiences with various businesses. These reviews are a goldmine of data that, when analysed correctly, can provide valuable insights into customer satisfaction. This project proposes a system to scrape reviews of Ice Cream shops in Toronto from Yelp, analyse the sentiment of these reviews, and visualize the results in a user-friendly manner. This project is intended for newcomers in Toronto and tourists paying a visit to the city which would help them get an easier and faster means of knowing the best in touch. This would also help the business owners in making their service and offerings better to get more revenue.

# 3. Project Description

**Data Collection**

We will scrape reviews from Yelp using a web scraping python script. The data will include the review text, rating, and other relevant information.

**Data Storage**

The scraped data will be stored in MongoDB, a NoSQL database. This will allow us to efficiently store and retrieve the data.

**Data Processing**

We will use RapidMiner to read the data from MongoDB using the Read MongoDB operator. The data will then be converted from JSON format to a format suitable for analysis. Data is also cleaned with the help of a python script to remove non alphanumeric values in the reviews.

**Sentiment Analysis**

We will use the Sentiment Analysis operator in RapidMiner to analyse the sentiment of the reviews. This will involve training a machine learning model on a labelled dataset and then applying this model to the Yelp reviews.

**Result Storage**

The sentiment values will be generated as attributes and stored in another collection in the MongoDB database.

**Data Visualization**

We will use Flask, a Python web framework, to create a web application for visualizing the sentiment analysis results.

# 4. Technology Stack

* Hardware requirement

MacBook Air or Pro / Manufacturers which support Windows operating system.

* Software Requirement

1. Python scripts: For scrapping reviews from yelp.
2. MongoDB: For storing and retrieving data.
3. RapidMiner: For data processing and sentiment analysis.
4. Flask: To display visualization on a Web application.

# 5. Project Scope

The scope of this project includes the collection, storage, processing, analysis, and visualization of Yelp reviews for sentiment analysis. The project will be considered successful when we can accurately determine the sentiment of a Yelp review and visualize this data in a user-friendly manner.

**In-Scope**

* Collection of Yelp reviews using a web scraping python script.
* Storage of the scraped data in MongoDB.
* Processing of the data using RapidMiner, including reading the data from MongoDB, converting it from JSON, and performing sentiment analysis.
* Storage of the sentiment analysis results in MongoDB.
* Visualization of the sentiment analysis results using a Flask application.

**Out-of-Scope**

* Collection of reviews from platforms other than Yelp.
* Use of databases other than MongoDB for data storage.
* Use of tools other than RapidMiner for data processing and sentiment analysis.
* Use of frameworks other than Flask for data visualization.

# 6. Implementation Plan

1. Set up the MongoDB database.
2. Develop the web scraping python script to collect Yelp reviews and Business Names.
3. Store the scraped data in MongoDB in a collection named Collection 3.
4. Develop the RapidMiner process for reading the data, converting it, clean and performing sentiment analysis.
5. Store the sentiment analysis results in MongoDB.
6. Develop the Flask application for data visualization.
7. Test the entire system and make necessary adjustments.
8. Launch the system and start collecting and analysing Yelp reviews.

# 7. Design Diagram

A screenshot of a computer game

Description automatically generated

# 8. Risk Assessment

* Potential risks include changes to the Yelp website that could break the web scraping script, and inaccuracies in the sentiment analysis results. We will mitigate these risks by regularly updating the web scraping python script and by continuously improving the sentiment analysis model.
* Yelp might change its website structure, which would require us to update our web scraping python script. This can be done by enabling logging on the script, which is a potential feature that could be added in the next development phase.

# 9. Cost Benefit Analysis

The main costs of this project will be the development and maintenance of the system. The benefits will be the valuable insights gained from the sentiment analysis, which can be used to improve customer satisfaction and ease customer selection of ice cream shops while they are visiting Toronto and folks who are new to the place.

# 10. Deliverables

1. A web scraping python script for collecting Yelp reviews.
2. A MongoDB database for storing the scraped data and the sentiment analysis results.
3. A RapidMiner process for reading the data from MongoDB, converting it, and performing sentiment analysis.
4. A Flask application for visualizing the sentiment analysis results.

# 11. Conclusion

This project gives us valuable insights into customer sentiment and help improve the service of ice cream shots. By investing in this project, we can increase customer satisfaction and increase revenue. This tool would be helpful for tourists and new comes in Toronto to visit ice cream shops in Toronto.

# 12. References and Citation

[Scrape Yelp Data using Python [Build Your Yelp Scraper] (scrapingdog.com)](https://www.scrapingdog.com/blog/how-to-scrape-yelp-reviews/)

[Beautiful Soup: Build a Web Scraper With Python – Real Python](https://realpython.com/beautiful-soup-web-scraper-python/)

[Flask - (Creating first simple application) - GeeksforGeeks](https://www.geeksforgeeks.org/flask-creating-first-simple-application/)

[How to build a web application using Flask and deploy it to the cloud (freecodecamp.org)](https://www.freecodecamp.org/news/how-to-build-a-web-application-using-flask-and-deploy-it-to-the-cloud-3551c985e492/)

[Build A Python Database With MongoDB | MongoDB](https://www.mongodb.com/languages/python)

[Connection Strings — MongoDB Manual](https://www.mongodb.com/docs/manual/reference/connection-string/)

[Can I use the connection string generated in the MongoDB Compass in Pymongo? - Stack Overflow](https://stackoverflow.com/questions/63568938/can-i-use-the-connection-string-generated-in-the-mongodb-compass-in-pymongo)

[Using MongoDB - RapidMiner Documentation](https://docs.rapidminer.com/9.4/studio/connect/database/mongodb.html)

[How to Integrate MongoDB Into RapidMiner - Zebra BI](https://zebrabi.com/guide/how-to-integrate-mongodb-into-rapidminer/)

[Putting data from MongoDB into a chart in python (using pandas dataframe) - Stack Overflow](https://stackoverflow.com/questions/70683894/putting-data-from-mongodb-into-a-chart-in-python-using-pandas-dataframe)

[Plot a pie chart in Python using Matplotlib - GeeksforGeeks](https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/)

[Add HTML and CSS in Flask Web Application - ThinkInfi](https://thinkinfi.com/flask-adding-html-and-css/)

[Flask/Python - handling dropdown to open different HTML pages - Stack Overflow](https://stackoverflow.com/questions/35264455/flask-python-handling-dropdown-to-open-different-html-pages)

[HTML : How to create dropdown menu from python list using Flask and HTML - YouTube](https://www.youtube.com/watch?v=qtde16rN27U)

[Aggregation Stages — MongoDB Manual](https://www.mongodb.com/docs/manual/reference/operator/aggregation-pipeline/)

[Query and Projection Operators — MongoDB Manual](https://www.mongodb.com/docs/manual/reference/operator/query/)

[What is the re.sub() function in Python? (educative.io)](https://www.educative.io/answers/what-is-the-resub-function-in-python)

[Python | Regular Expressions | re.sub() | Codecademy](https://www.codecademy.com/resources/docs/python/regex/sub)