# Configuration Manual

MSc Research Project
Data Analytics

## Anil Prajapati

Student ID: X23167297

School of Computing
National College of Ireland

Supervisor:     Musfira Jilani

# National College of Ireland
## Project Submission Sheet
### School of Computing

| | |
|---|---|
| **Student Name:** | Anil Prajapati |
| **Student ID:** | X23167297 |
| **Programme:** | Data Analytics |
| **Year:** | 2023 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Musfira Jilani |
| **Submission Due Date:** | 12/12/2024 |
| **Project Title:** | Customer Sentiment Analysis for Service Issue Detection Using Negative Sentiment |
| **Word Count:** | 240 |
| **Page Count:** | 5 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | |
| **Date:** | 12th December 2024 |

## PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | ☐ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | ☐ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | ☐ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Configuration Manual

## Anil Prajapati
## X23167297

# 1 System Details

The project was executed on a personal windows laptop with an x86 processor and dedicated GPU. Although not require but is recommended to have a dedicated GPU.

| OS | Windows 11 |
|---|---|
| OS Version | 24H2 |
| System RAM | 16GB |
| GPU | RTX3050 4GB |
| Processor | AMD Ryzen 7 5800H |
| Python version | 3.10.12 |

Table 1: Your table caption here

# 2 Download ollama

Step 1 - Download ollama from the below link
https://ollama.com/download
Step 2 - Open the commanda prompt and run the below command to downland and run llama model

ollama run llama3.3 ollama (n.d.)



```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\anilp> ollama run llama3.3
```

Figure 1: ollama

# 3 Coding

This section we will run the code

## 3.1   Downlaod data file

We will run the below cell to downlaod the data file and save as an excel file.

```python
import steamreviews
import pandas as pd
from transformers import pipeline
import torch
from datasets import Dataset


app_id = 570

#app_id = [730, 570, 1172470, 271590, 1091500, 1174180, 1086940, 1245620, 1716740, 990080]
games = { …

review_dict, query_count = steamreviews.download_reviews_for_app_id(app_id)
reviews_data = review_dict['reviews']

review_list = []
for review_id, review in reviews_data.items():
    # Filter for English reviews only
    if review['language'] == 'english':
        review_list.append({
            'review_id': review_id,
            'language': review['language'],
            'review': review['review'],
            'timestamp_created': review['timestamp_created'],
            'voted_up': review['voted_up'],
            'votes_up': review['votes_up'],
            'votes_funny': review['votes_funny'],
            'weighted_vote_score': review['weighted_vote_score'],
            'comment_count': review['comment_count']
        })

df = pd.DataFrame(review_list)
df.to_excel('D:\\Study\\Research Project\\dota2_steam_reviews.xlsx', index=False)
```

## 3.2 Sentiemnt Analysis

Run the below code to perform the sentiment analysis and save the sentiment data as an excel file.

# Sentiment Analysis

using Bert, distilbert and Vader

```python
max_length = 512
transformer_model_name = "distilbert-base-uncased-finetuned-sst-2-english"
sentiment_pipeline = pipeline("sentiment-analysis", device=0, truncation=True)
sentiment_pipeline_alt = pipeline("sentiment-analysis", model=transformer_model_name, device=0)
vader_analyzer = SentimentIntensityAnalyzer()


df = pd.read_excel('D:\\Study\\Research Project\\dota2_steam_reviews.xlsx')
df['review'] = df['review'].fillna('')
df['review'] = df['review'].apply(lambda x: x[:max_length])  # Truncate long reviews
dataset = Dataset.from_pandas(df)


def analyze_batch(batch):
    sentiments = sentiment_pipeline(batch['review'])
    return {'sentiment': [result['label'] for result in sentiments]}

def analyze_batch_alt(batch):
    sentiments = sentiment_pipeline_alt(batch['review'])
    return {'sentiment_alt': [result['label'] for result in sentiments]}
```

```python
def analyze_batch_alt(batch):
    sentiments = sentiment_pipeline_alt(batch['review'])
    return {'sentiment_alt': [result['label'] for result in sentiments]}

def analyze_batch_vader(batch):
    vader_scores = [vader_analyzer.polarity_scores(text)['compound'] for text in batch['review']]
    vader_sentiments = [
        "Positive" if score > 0.05 else "Negative" if score < -0.05 else "Neutral"
        for score in vader_scores
    ]
    return {'vader_sentiment': vader_sentiments}


dataset = dataset.map(analyze_batch, batched=True, batch_size=64)          # Model 1 Bert
dataset = dataset.map(analyze_batch_alt, batched=True, batch_size=64)   # Model 2 distilbert
dataset = dataset.map(analyze_batch_vader, batched=True, batch_size=64) # VADER


df_lab = dataset.to_pandas()


output_file = 'D:\\Study\\Research Project\\dota2_comparison_reviews_with_vader.xlsx'
df_lab.to_excel(output_file, index=False)
print(f"Results saved to {output_file}")


agreement_transformers = df_lab['sentiment'] == df_lab['sentiment_alt']
agreement_transformer_vader = df_lab['sentiment'] == df_lab['vader_sentiment']
agreement_alt_vader = df_lab['sentiment_alt'] == df_lab['vader_sentiment']

print(f"Agreement between Transformer Model 1 and 2: {agreement_transformers.mean():.2%}")
print(f"Agreement between Transformer Model 1 and VADER: {agreement_transformer_vader.mean():.2%}")
print(f"Agreement between Transformer Model 2 and VADER: {agreement_alt_vader.mean():.2%}")
```

```python
plt.figure(figsize=(12, 6))


sns.countplot(data=df_lab, x='sentiment', label='Model 1', color='blue', alpha=0.6)


sns.countplot(data=df_lab, x='sentiment_alt', label='Model 2', color='orange', alpha=0.6)
# VADER
sns.countplot(data=df_lab, x='vader_sentiment', label='VADER', color='green', alpha=0.6)

plt.legend()
plt.title('Sentiment Distribution Comparison')
plt.show()
```

## 3.3    Llama for issue highlighting

We will run this cell to query llama 3.2 to label the issue

4

```
from ollama import Client
import pandas as pd

client = Client(
  host='http://localhost:11434',
  headers={'x-some-header': 'some-value'}
)

file = open("D:\\Study\\Research Project\\res.txt", "a", encoding="utf-8")
file.truncate(0)

def func(data):
  query = '[crashes, match making, fps drop, performance, no-issue] out of these common issues in games, which issue can you detect in the following comment ' + '"'+ str(data) +'".
  response = client.chat(model='llama3.2', messages=[
    {
      'role': 'user',
      'content': query,
    },
  ])
  file.write(response.message.content + "\n")
  return response.message.content

file_path = "D:\\Study\\Research Project\\dota2_steam_reviews_classified.xlsx"


df = pd.read_excel(file_path, sheet_name=0, header=0, index_col=None)
#df = df[0:10]
df["issue"] = df["review"].apply(func)

file.close()
output_file = 'D:\\Study\\Research Project\\dota2_steam_reviews_classified_issue.xlsx'
df.to_excel(output_file, index=False)
```

# References

ollama (n.d.). ollama.
URL: *https://ollama.com/library/llama3.3*