National College of Ireland

Configuration Manual: Detecting Customer Dissatisfaction in Support Chats Using AI-Based Sentiment Analysis

MSc Research Project
MSc Artificial Intelligence

Diego Lemos
Student ID: x20204787

School of Computing
National College of Ireland

Supervisor: Dr. Devanshu Anand

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Diego Lemos
Student ID:	x20204787
Programme:	MSc Artificial Intelligence
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr. Devanshu Anand
Submission Due Date:	11/08/2025
Project Title:	Configuration Manual: Detecting Customer Dissatisfaction in Support Chats Using AI-Based Sentiment Analysis
Word Count:	460
Page Count:	3

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:	Diego Lemos
Date:	9th August 2025

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: Detecting Customer Dissatisfaction in Support Chats Using AI-Based Sentiment Analysis

Diego Lemos
x20204787

1. System Requirements

1.1 Hardware

- CPU: Any 4-core+ processor (e.g., Intel i5/i7, Apple M1, Ryzen)
- RAM: 8 GB minimum (16 GB recommended)
- Storage: 2–5 GB of free disk space
- GPU: Optional (for faster BiLSTM training)

1.2 Software

- Operating System: macOS / Windows / Linux
- Python: Version 3.9 or higher
- Jupyter Notebook or VS Code
- Required packages listed in `requirements.txt`:
  - numpy, pandas, matplotlib, seaborn
  - scikit-learn, torch, transformers

2. Environment Setup

1. Clone the project from GitHub or unzip the provided artefact.
2. Create and activate a virtual environment:

```
python -m venv venv
source venv/bin/activate (macOS/Linux)
venv\Scripts\activate (Windows)
```

3. Install dependencies:

```
pip install -r requirements.txt
```

3. Dataset Preparation

The dataset used was the Twitter Customer Support Dataset, which is publicly available on Kaggle at: `https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter/data`.

Due to file size restrictions the dataset is not included in this submission. It should be downloaded from the provided link and placed in the `data/raw/` directory.

The preprocessing script (`01_preprocessing.ipynb`) will:

- Clean and lowercase text
- Label sentiments (VADER, BERT, RoBERTa)
- Tokenize and split into train/test sets

4. Configuration and Parameters

You will be able to configure the model type, embeddings, and other parameters inside each notebook. For example:

- Model type: SVM, Logistic Regression, Random Forest and BiLSTM
- Embedding: TF-IDF, PCA, Token Embedding
- Split: 70/30, 75/25, 80/20
- Epochs, batch size: Defined in BiLSTM notebook

5. Execution Instructions

Traditional ML Models:

`notebooks/02_model_training_ML.ipynb`

BiLSTM:

`notebooks/03_bilstm_training.ipynb`

Figures are exported to results as JSON files, and CSV files are also exported to `results/`.

6. Logging and Output

Outputs includes:

- CSV result files in `results/`
- Confusion matrices
- Model comparison charts
- Summary metrics in printed JSON format

7. Reproducibility

To ensure consistent results, we use:

- `random.seed(42)` in Python
- `torch.manual_seed(42)` for BiLSTM

Due to some randomness in transformer-based models slight variations may still occur.