📘 Automated IT Service Ticket Classifier - Documentation

# 1. Project Overview

This project automates the classification of IT service tickets into predefined categories using Natural Language Processing (NLP) and Deep Learning. It consists of two main parts: a training script for preparing and training the model, and a Streamlit application for real-time deployment.

# 2. Workflow

Dataset (CSV) → Preprocessing (Cleaning → Tokenization → Padding) → LSTM Model Training → Save Artifacts (.h5, .pkl) → Streamlit App Deployment → User Prediction

# 3. Training Script (train\_ticket\_classifier.py)

* Load dataset (CSV containing ticket descriptions and categories)
* Encode labels into one-hot vectors and save category mapping
* Clean text (lowercase, remove punctuation)
* Tokenize and pad text sequences
* Split data into training and validation sets
* Build LSTM model (Embedding → LSTM → Dense → Dropout → Softmax)
* Train model and save artifacts: ticket\_classifier.h5, tokenizer.pkl, category\_mapping.pkl

# 4. Streamlit App (app.py)

* Loads model, tokenizer, and category mapping
* Provides text input for ticket description
* Cleans and preprocesses input text
* Uses trained model to predict ticket category
* Displays predicted category and confidence score in Streamlit interface

# 5. Key Files & Outputs

|  |  |
| --- | --- |
| File | Purpose |
| it\_service\_tickets.csv | Dataset with ticket descriptions and categories |
| train\_ticket\_classifier.py | Training pipeline for LSTM model |
| ticket\_classifier.h5 | Trained model |
| tokenizer.pkl | Tokenizer object for preprocessing |
| category\_mapping.pkl | Mapping between categories and indices |
| app.py | Streamlit app for deployment |

# 6. How to Run

Training: `python train\_ticket\_classifier.py`

Run Streamlit App: `streamlit run app.py`

# Dataset Details:

Link: <https://www.kaggle.com/datasets/adisongoh/it-service-ticket-classification-dataset?utm_source=chatgpt.com>

You can use the IT Service Ticket Classification Dataset from Kaggle.

This dataset contains:

* **47,837 rows** of ticket data.
* **2 features**:
  + Document: The ticket text/content.
  + Topic\_group: The ticket category.

Categories include: "Hardware," "Software," "Network," "Access," "Service," and "Other."

# 7. Future Improvements

* Train with more epochs for higher accuracy
* Experiment with Bidirectional LSTM, GRU, or Transformers
* Add visualizations (word clouds, class distribution) in Streamlit
* Support multi-label classification for complex tickets