Banking Support Intent Classification using NLP (Banking77)

A transformer-based classifier that routes banking support queries into 77 fine-grained intents — the routing brain behind chatbots and ticket triage, trained on PolyAI's Banking77 benchmark.

1. Introduction / Background

Every bank's chatbot and support desk faces the same first problem: given a customer's message — 'I lost my debit card yesterday' — which of dozens of possible intents does it express? Get that routing wrong and the customer is sent down the wrong flow.

This project fine-tunes DistilBERT on Banking77 (13,083 real-style banking queries, 77 fine-grained intents) — a transformer small enough to train on a single GPU yet strong enough to beat classical baselines convincingly — and evaluates with macro F1 across all 77 intents. An interactive demo shows the routing decision with confidence scores and a suggested response.

2. Problem Statement

Misrouted support queries waste agent time and frustrate customers; keyword rules break on paraphrase ('debit card gone' vs 'lost my card'). A learned intent classifier generalizes across phrasing, but students need to see the full pipeline — tokenization, fine-tuning, honest multi-class evaluation — not just an API call.

3. Objectives

- Fine-tune DistilBERT to classify banking queries into 77 intents.
- Evaluate with accuracy, macro F1 and per-intent precision/recall on a held-out test set.
- Analyze the most-confused intent pairs for the report.
- Build an interactive demo showing the predicted intent with evidence and a suggested response.

4. Proposed Methodology

Banking77 queries are split into train/validation/test sets and tokenized with the DistilBERT WordPiece tokenizer (max length 64) with padding and attention masks. distilbert-base-uncased receives a classification head (dropout → dense-77 → softmax) and is fine-tuned with AdamW (lr 2e-5) for up to 5 epochs (batch 32), with early stopping on validation macro F1. The held-out test set is evaluated once, with per-intent metrics and a confusion analysis of the hardest intent pairs.

5. System Architecture / Working

- Input: customer query → WordPiece tokens (max 64) + attention mask.
- Encoder: DistilBERT, 6 layers, 768-dim [CLS] representation.

- Head: dropout → dense 77 → softmax (intent probabilities).
- Demo: typed query → representative-intent scoring → top intent + confidence bars + suggested response.

6. Hardware / Software Requirements

- Python 3, PyTorch, Hugging Face Transformers, scikit-learn, Jupyter Notebook.
- A GPU is recommended for fine-tuning (CPU training is possible but slow).
- The web demo runs offline in any modern browser after download.

7. Expected Outcomes

- A fine-tuned DistilBERT reaching the design targets of ≈ 93% test accuracy and ≈ 0.92 macro F1, reported honestly after the training run.
- An interactive intent-classification demo with suggested responses.
- A complete report, PPT and viva Q&A; covering transformers and fine-tuning.

8. Applications

- Chatbot intent routing and support-ticket triage demonstration.
- Teaching transformer fine-tuning on a real benchmark.
- A template for intent classification in other domains (telecom, e-commerce).

9. Future Scope

- Out-of-scope detection for queries outside the 77 intents.
- Multilingual intent classification.
- Distillation to a tiny model for on-device inference.

10. Conclusion

The project delivers the complete modern NLP loop — benchmark data, transformer fine-tuning, honest multi-class evaluation and a working demo — showing exactly how a chatbot decides what a customer wants.