

PROJECT ABSTRACT

Smart Email Triage Dashboard with Urgency Classification

A DistilBERT classifier trained on Enron emails triages an inbox by urgency and topic into an explainable priority queue, with suggested next actions.

01. Title

Smart Email Triage Dashboard with Urgency Classification

02. Introduction / Background

Professionals receive dozens of emails daily and the important ones hide among newsletters and FYIs; chronological inboxes treat a server-outage alert and a lunch invitation identically. This project builds an urgency-aware triage layer: a DistilBERT classifier fine-tuned on the Enron email corpus (half a million real workplace emails, labeled subset for urgency and topic) scores every message, a transparent rule layer adds sender-priority and deadline signals as auditable rationale, and the inbox re-renders as a priority queue with suggested next actions.

03. Problem Statement

Time-sensitive email gets buried chronologically, causing missed deadlines and slow incident response, while existing priority features are opaque. The project addresses this with a documented two-stage pipeline — transformer classification plus transparent signal rules — where every urgency score carries its rationale and the queue ranking is fully inspectable.

04. Objectives

- Fine-tune DistilBERT with two heads: 4-level urgency and 5-category topic classification.
- Create documented urgency/topic labeling rules and a labeled Enron subset.
- Build a transparent signal layer (sender priority, deadlines, escalation keywords) with per-email rationale.
- Render the inbox as a ranked priority queue with category filters.
- Generate template-based suggested actions per (urgency, category) pair.
- Evaluate with macro-F1, per-class precision/recall and confusion matrix on a held-out set; ship a live demo on a realistic 24-message inbox.

05. Proposed Methodology

Enron emails are parsed, reply chains stripped, and a documented subset labeled for urgency (Critical/High/Normal/Low) and topic (Finance/Operations/HR/External/Personal) using published labeling rules. DistilBERT-base is fine-tuned with two classification heads. At inference, classifier scores combine with rule-layer signals (VIP sender lists, deadline mentions, escalation keywords); messages sort by adjusted urgency then recency. The Flask demo triages the preloaded inbox instantly with urgency/topic filters and a queue view.

06. System Architecture / Working

Raw email → parsing/cleaning → DistilBERT urgency+topic heads → rule-layer adjustment with rationale capture → priority queue ranking → template action suggester → demo UI (triaged inbox, email detail with rationale, priority queue with scores). Suggested actions are text only — nothing is ever sent automatically.

07. Hardware / Software Requirements

- Python 3.11, PyTorch, Hugging Face Transformers (DistilBERT), scikit-learn, pandas
- Enron Email Dataset (~500k emails; documented labeled subset)
- Flask + HTML/CSS/JS demo
- CPU suffices for inference (<50 ms/email); GPU speeds fine-tuning

08. Expected Outcomes

A triage dashboard that surfaces critical mail above promotions; macro-F1 and per-class metrics on the held-out labeled set reported as achieved figures after training (design targets until then); a demo inbox where every score is explainable.

09. Applications

- Inbox prioritization for busy professionals and support teams
- Incident-mail surfacing for on-call rotations
- Teaching text classification with a real corpus and honest labeling
- Executive-assistant style daily priority digest

10. Future Scope

- On-device processing design for private-mail privacy
- Thread-level triage (whole conversation, not single messages)
- Calendar-aware urgency (deadlines from events)
- Active learning: urgency labels refined from user re-ranking