

Implementation – Email Classification using Transformers

Tim Terbach

June 2025

Contents

1	Einleitung	2
1.1	Projekt	2
2	Projektstruktur	2
2.1	Frontend	2
2.2	Backend	3
2.2.1	RestAPI	3
2.2.2	Docker	3
3	Trainingsdaten	4
3.1	E-Mail Generierung mit der ChatGPT API	4
3.1.1	User Prompt	5
3.1.2	Generierter Datensatz	6
4	Fine Tuning	6
4.1	Auswahl des BERT Modells	6
4.2	Modell auf Training vorbereiten	7
5	Auswertung	7

1 Einleitung

Nachdem in der theory.pdf verallgemeinert die theoretischen Grundbausteine dieses Projektes erklärt wurden, beschäftigt sich diese Datei nun mit der konkreten Anwendung dieses Wissens zum Lösen eines praktischen Problems.

1.1 Projekt

In diesem Projekt wollen wir das Postfach eines Versicherungsunternehmens simulieren. Genauer sollen dabei eingehende E-Mails in fünf verschiedene Kategorien eingeordnet werden:

1. Kfz-Schaden
 - Unfall, Blebschaden, Wildunfall
2. Hausrat-Schaden
 - Einbruch, Wasserschaden, Brandschaden
3. Haftpflicht
 - versehentliche Beschädigung fremder Dinge
4. Reiseschaden
 - Gepäckverlust, Stornierung, Krankheit im Urlaub
5. Tierkrankheit
 - OP beim Haustier, Tierarztrechnung

Anschließend können diese E-Mails dann den richtigen Sachbearbeitern zugeordnet werden. Da es sich hierbei um ein Demoprojekt handelt, werden diese zu klassifizierenden E-Mails eine verhältnismäßig einfache Struktur haben und sich inhaltlich pro Label auf die 2-3 vordefinierte Themen fokussieren.

2 Projektstruktur

An dieser Stelle gehen wir kurz auf die allgemeine Projektstruktur ein, sowie die gewählten Architekturen und Technologien.

```
VectorMail/
api/          # Spring Boot backend (REST API)
frontend/     # Vue.js frontend (Vite)
python-service/ # Python-based classification module (Flask)
docs/         # Theory and implementation details
docker-compose.yml # Launches the Flask engine container
README.md     # Project overview
```

2.1 Frontend

Das Frontend wird mit Vue.js entwickelt und soll sich durch ein verspieltes, dennoch übersichtliches und intuitives Design auszeichnen.

Beim Styling habe ich mich für den BEM-Syntax entschieden, um ein besseres Verständnis für CSS-Prinzipien zu erlangen und gleichzeitig eine professionelle Benennung meiner CSS Elemente zu garantieren.

Für mehr Informationen betrachten Sie die README.md des frontend-Verzeichnisses.

2.2 Backend

Das Backend wird als REST-API umgesetzt und übernimmt das Klassifizieren der E-Mails mithilfe unseres trainierten LLMs. Es wird in einem Docker-Container mit Flask betrieben, um eine isolierte, skalierbare und leicht wartbare Architektur zu gewährleisten. Python wurde als Programmiersprache für das Training von BERT gewählt, da es weit verbreitet in KI-Anwendungen und Algorithmik ist und ich zudem meine Python Fähigkeiten ausbauen möchte.

2.2.1 RestAPI

Die REST-API dient als Schnittstelle zwischen dem Client und dem Backend. Der erste implementierte Endpunkt ist eine POST-Request, die die Wahrscheinlichkeit der definierten Labels zurückgibt. Dabei dient das Backend (Spring Boot) als Vermittler zwischen Client und unserem Docker Container, welcher das Python Skript beinhaltet. Dort wird die eigentliche Klassifizierung durchgeführt.



Figure 1: Symbolische Darstellung der Kommunikation

2.2.2 Docker

Um eine modulare und leicht wartbare Architektur sicherzustellen, wird das Python-Programm, wie bereits erwähnt, in einem separaten Docker-Container betrieben.

`/python-service` enthält alle notwendigen Dateien für den Python-Dienst, darunter:

- `requirements.txt` zur Definition der benötigten Python-Abhängigkeiten
- `Dockerfile` zur Containerisierung des Backends
- `TODO...`

Zunächst habe ich mich auch in diesem Projekt für `'python:3.10'` als Image entschieden, da es bereits in zuverigen Projekten Probleme mit dem schlankeren `python:3.10-slim` gab.

Um dafür zu sorgen, dass sich der Dockercontainer selbstständig aktualisiert, Änderungen also selbstständig übernommen werden, ohne den Container mit `docker-compose up` neu bauen zu müssen, habe ich mich hier dafür entschieden diese Funktion mithilfe von Volumes zu ermöglichen.

Durch die Erweiterung der `docker-compose.yml` um

```
volumes:
- ./python-service:/app
```

erkennt der Dockercontainer nun, falls Änderungen innerhalb des “python-service“ Verzeichnis vorgenommen werden und startet den Container automatisch neu.

3 Trainingsdaten

Um BERT zu Fine Tunen brauchen wir zunächst Trainingsdaten. Zum Start wollen wir dabei mit etwa 200 E-Mails pro Label beginnen. Da es keine passende Datenbank gibt, die E-Mails zu diesen Kategorien liefert, habe ich mich an dieser Stelle dazu entschieden diese von einem weiteren, generativen LLM erzeugen zu lassen, in diesem Fall ChatGPT. Bei einer Menge von insgesamt 1000 E-Mails müssen wir dabei jedoch einen anderen Ansatz wählen, als einfach ChatGPT.com zu öffnen und nach 1000 E-Mails zu diesen Labels zu fragen. Wir nutzen hier die sehr einfache API von ChatGPT um per Python-script diese E-Mails zu generieren. Obwohl dieser Schritt bei der Planung des Projektes zunächst nicht angedacht war, dient er als perfekte Erweiterung des Projektes, da wir auch hier weitere Konzepte aus der theory.pdf aufgreifen.

3.1 E-Mail Generierung mit der ChatGPT API

Betrachten wir zunächst mal das Python-script welches wir zu Generierung verwenden werden, genauer die `generate_email()` Funktion, welche iterativ ausgeführt wird und somit immer wieder neue E-Mails liefert.

```
def generate_email():
    response = client.chat.completions.create(
        model="gpt-3.5-turbo",
        messages=[
            {"role": "system", "content": "Du bist ein Generator für realistische, formelle E-Mails an eine Versicherung. Gib nur JSON zurück."},
            {"role": "user", "content": base_prompt}
        ],
        temperature=0.8
    )
    text = response.choices[0].message.content
    try:
        return json.loads(text)
    except:
        print("Fehler beim Parsen der Antwort:")
        print(text)
        return None
```

Diese Funktion ist relativ intuitiv und bedarf keiner großartiger Erklärung, interessant sind jedoch vor allem die Paramter

- `temperature=0.8`
- `messages= ["role": "system", "content": "Du bist ein Generator für realistische, formelle E-Mails an eine Versicherung. Gib nur JSON zurück.", "role": "user", "content": base_prompt],`

Die **temprature** sagt hierbei aus, wie zufällig das Modell bei der Tokengenerierung vorgehen soll. Bei einer Temprature von 0.0 würde das Modell immer nur den warhscheinlichsten nächsten Token wählen, bei einem gleichbleibenden Prompt würde somit auch der Output bzw. die Sequenz generierter Tokens gleich bleiben, das wollen wir hier natürlich verhindern. Mit einem temprature-Wert von 0.8 sagen wir dem Modell

also, dass aus auch Tokens wählen darf, welche nicht am wahrscheinlichsten von allen Tokens als nächste vorkommen, um eine hohe Variation an verschiedenen E-Mails zu erzeugen.

Ein weiterer interessanter Punkt sind die übergebenen **messages**. Mit **messages** wird jeweils eine Rolle sowie ein Prompt (Content) übergeben. Dabei legen wir mit **role="system"** fest, dass es sich um einen System Prompt handelt. Hier wird dem Modell kurz und knapp gesagt, wie es sich verhalten soll, in unserem Fall legen wir hier vor allem fest, dass das Modell nur mit einer JSON antworten soll, da wir mehr Output hier nicht brauchen, bzw. dieser zusätzliche Output mit mehr Tokens und somit höheren Kosten verbunden wäre.

Anschließend übergeben wir den eigentlicher Nutzer-Prompt, dementsprechend mit **role="user"**. Dieser Prompt bezieht sich vor allem auf ein paar inhaltliche Richtlinien und Beispiele der E-Mails für die verschiedenen Labels. Den kompletten User-Prompt hat dabei ChatGPT selbst auf Basis dieser Richtlinien und Beispiele erstellt, mit dem Ziel ein weiteres LLM dazu anzuweisen, einen für diese Aufgabe optimalen Datensatz zu erzeugen, diesen können Sie in dem folgenden Abschnitt betrachten.

3.1.1 User Prompt

Bitte generiere jeweils genau **eine E-Mail** für jedes Label im folgenden JSON-Format:

```
{
  "label": "<schadenstyp1>",
  "text": "<realistische E-Mail in deutscher Sprache>"
}
{
  "label": "<schadenstyp2>",
  "text": "<realistische E-Mail in deutscher Sprache>"
}
{
  "label": "<schadenstyp3>",
  "text": "<realistische E-Mail in deutscher Sprache>"
}
{
  "label": "<schadenstyp4>",
  "text": "<realistische E-Mail in deutscher Sprache>"
}
{
  "label": "<schadenstyp5>",
  "text": "<realistische E-Mail in deutscher Sprache>"
}
```

Regeln:

1. Die E-Mail soll **formell, aber alltagsnah** klingen – wie eine echte Nachricht von Privatpersonen oder Kunden.
2. Der E-Mailtext soll **3–4 Sätze** lang sein, in natürlichem Deutsch.
3. Verwende **keine sensiblen Daten** wie echte Adressen, Telefonnummern oder reale Versicherungsnummern.
4. Die E-Mail soll **thematisch zum jeweiligen Schadenstyp passen**, ohne zu ausschweifend oder künstlich zu wirken.
5. Verwende **normale, höfliche Sprache**, gerne mit kleinen individuellen Variationen im Ausdruck.
6. Gib **nur die Keys label und text** zurück – keine Metadaten, Überschriften oder Kommentare.

Mögliche Labels (zufällig auswählen):

- "kfz-schaden" → z. B. Unfall, Blechschaden, Wildunfall
- "hausrat-schaden" → z. B. Einbruch, Wasserschaden, Brandschaden
- "haftpflichtschaden" → z. B. versehentliche Beschädigung fremder Dinge
- "reiseschaden" → z. B. Gepäckverlust, Stornierung, Krankheit im Urlaub
- "tierkrankheit" → z. B. OP beim Haustier, Tierarztrechnung

Thematische Richtlinien für jede Kategorie:

Kfz-Schaden:

- Betreff: Schäden an Autos oder Motorrädern
- Typische Begriffe: Unfall“, Fahrzeug“, Blechschaden“, Wild“, Polizei“

Hausrat-Schaden:

- Betreff: Schäden am Inventar in Wohnung oder Haus
- Typische Begriffe: Einbruch“, Rohrbruch“, Fernseher“, gestohlen“

Haftpflichtschaden:

- Betreff: Absender hat Dritten Schaden zugefügt
- Typische Begriffe: versehentlich“, mein Kind“, beschädigt“, Nachbar“

Reiseschaden:

- Betreff: Probleme im Urlaub, z. B. Verlust, Krankheit
- Typische Begriffe: Koffer“, Flug“, Storno“, Rechnung“, Ausland“

Tierkrankheit:

- Betreff: Tierarztkosten, Behandlungen von Haustieren
- Typische Begriffe: Hund“, Katze“, OP“, Tierarzt“, Rechnung“

3.1.2 Generierter Datensatz

TODO

4 Fine Tuning

Nachdem wir nun also unsere E-Mails unter anderem mit Hilfe von System Prompting generiert haben, wollen wir diese Daten für das Fine Tuning eines BERT Modells nutzen.

4.1 Auswahl des BERT Modells

Zunächst müssen wir uns natürlich die Frage stellen, welches BERT Modell wir für unser Projekt überhaupt benutzen. Seit seiner Veröffentlichung 2018 gibt es viele abgewandelte oder optimierte Versionen dieses LLMs, bekannte Beispiele sind dabei:

- **DistilBERT** (HuggingFace): Eine kleinere und schnellere Version von BERT, die durch Destillation trainiert wurde und etwa 95% der Performance des originalen BERT beibehält.
- **RoBERTa** (Meta): Eine robust optimierte Variante von BERT, die durch längeres Training auf mehr Daten und ohne die Next-Sentence Prediction Aufgabe eine verbesserte Leistung erzielt.

- **ALBERT** (Google): Eine "Lite BERT"-Version, die durch Parameter-Sharing und Embedding-Faktorisierung die Modellgröße drastisch reduziert, während die Performance erhalten bleibt.

Neben Variationen, welche Training und Parameteranzahl optimieren, gibt es auch Varianten, welche auf spezifischen Sprachen trainiert wurden, darunter auch **bert-base-german-cased**. Dieses Modell bietet sich für unser Projekt besonders gut an, da es durch seinen deutschen Trainingssatz besonders gut darin ist, den semantischen und kontextuellen Inhalt einer deutschen Textsequenz zu ergreifen und verstehen, genau das, was wir für dieses Projekt benötigen.

4.2 Modell auf Training vorbereiten

5 Auswertung